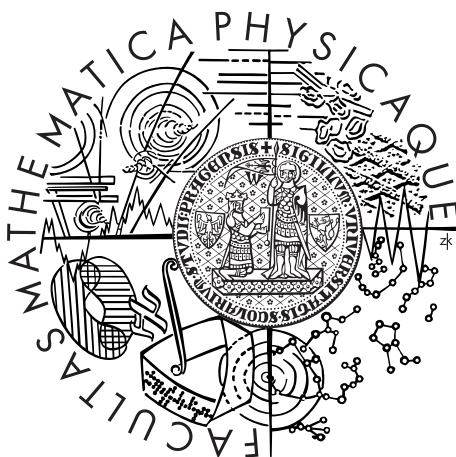


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Ivana Meňhartová

Metody dynamické analýzy složení portfolia

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Mgr. Tomáš Hanzák

Studijní program: Matematika

Studijní obor: MFAPM

Praha 2012

Rada by som poďakovala všetkým, ktorí mi pomohli, alebo ma akokoľvek podpo-
rili pri písaní tejto diplomovej práce. V prvom rade ďakujem môjmu vedúcemu
Mgr. Tomášovi Hanzákovi, za množstvo pripomienok a nápadov a za konzultácie
s možnosťou diskutovať o problémoch, ktoré sa pri písaní práce vyskytli. Ďalej by
som chcela poďakovať Lubomírovi Šerému za pomoc s programom Mathematica.

Prehlasujem, že som túto diplomovú prácu vypracovala samostatne a výhradne s použitím citovaných prameňov, literatúry a ďalších odborných zdrojov.

Beriem na vedomie, že sa na moju prácu vzťahujú práva a povinnosti vyplývajúce zo zákona č. 121/2000 Zb., autorského zákona v platnom znení, najmä skutočnosť, že Univerzita Karlova v Praze má právo na uzavretie licenčnej zmluvy o využití tejto práce ako školského diela podľa §60 odst. 1 autorského zákona.

V Prahe dňa

Podpis autora

Název práce: Metody dynamické analýzy složení portfolia

Autor: Ivana Meňhartová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Mgr. Tomáš Hanzák, KPMS, MFF UK

Abstrakt: V predloženej práci študujeme metódy používané k dynamickej analýze zloženia portfólia na základe jeho výnosov. Práca sa zameriava na Kalmanov filter a lokálne váženú regresiu, ako základné metódy dynamickej analýzy. Podrobne popisuje teóriu k týmto metódam, ich spôsob použitia a diskutuje ich vhodné nastavenie. V práci ďalej uvádzame praktické aplikácie oboch metód na umelo generovaných dátach, ale aj na reálnych dátach indexu pražskej burzy. Na umelých dátach rôznych typov tiež skúmame fungovanie modelu Kalmanovho filtra v prípade porušenia jeho predpokladov. V práci ďalej uvádzame pojem multikolinearity ako možnú komplikáciu v reálnych dátach. V závere práce porovnávame výsledky a využitie oboch metód a uvádzame možnosti rozšírenia Kalmanovho filtra pomocou projekcie odhadov a pomocou zavedenia tzv. CUSUM testov (testy detekcie zmien).

Klíčová slova: Kalmanov filter, lokálne vážená regresia, multikolinearita, CUSUM test

Title: Methods of dynamical analysis of portfolio composition

Author: Ivana Meňhartová

Department: Department of Probability and Mathematical Statistics

Supervisor: Mgr. Tomáš Hanzák, KPMS, MFF UK

Abstract: In the presented thesis we study methods used for dynamic analysis of portfolio based on it's revenues. The thesis focuses on Kalman filter and locally weighted regression as two basic methods for dynamic analysis. It describes in detail theory for these methods as well as their utilization and it discusses their proper settings. Practical applications of both methods on artificial data and real data from Prague stock-exchange are presented. Using artificial data we demonstrate practical importance of Kalman filter's assumptions. Afterwards we introduce term multicollinearity as a possible complication to real data applications. At the end of the thesis we compare results and usage of both methods and we introduce possibility of enhancing Kalman filter by projection of estimations or by CUSUM tests (change detection tests).

Keywords: Kalman filter, locally weighted regression, multicollinearity, CUSUM test

Obsah

Úvod	3
1 Základné informácie	4
1.1 Portfólio	4
1.2 Analýza portfólia	4
1.3 Stavové modelovanie	5
2 Kalmanov filter	6
2.1 Zavedenie Kalmanovho filtra	6
2.2 Počiatočné podmienky	8
2.3 Kalmanov filter pre portfólio	8
3 Praktické aplikácie Kalmanovho filtra	10
3.1 Výpočtové aspekty	10
3.2 Ideálny prípad	11
3.3 Nesplnenie predpokladov filtra	14
3.3.1 Závislosť náhodnej prechádzky inovácií a pozorovacieho šumu	14
3.3.2 Porušenie normality šumov	16
3.3.3 Šumy so skokmi	18
3.4 Príklad portfólia s reálnymi výnosmi	20
3.5 Vplyvy na konvergenciu	23
3.6 Reálny príklad PX indexu	30
4 Rolovacia regresia	33
4.1 Teoretický základ	33
4.2 Lokálne vážená regresia	33
4.3 Praktické aplikácie	34
4.3.1 Portfólio s umelo generovanými výnosmi	34
4.3.2 Portfólio s reálnymi výnosmi	36
4.4 Optimalizácia lokálne váženej regresie	40
4.5 Praktické aplikácie - zložitejšie príklady	41
5 Multikolinearita v dátach	44
5.1 Detekcia multikolinearity	44
5.2 Multikolinearita v cenách akcií	45
6 Porovnanie metód	46
6.1 Porovnanie na umelých a poloumelých dátach	46
6.2 Reálny príklad PX indexu	50
6.3 Zhrnutie poznatkov	51
7 Vylepšenie metód	53
7.1 Rozšírenie Kalmanovho filtra	53
7.1.1 Projekcia odhadov	54
7.1.2 CUSUM Kalman filter	56

Záver	60
Zoznam použitej literatúry	61
Prílohy	62

Úvod

Témou tejto diplomovej práce sú metódy použiteľné k dynamickej analýze zloženia portfólia - Kalmanov filter a lokálne vážená regresia. Práca nadväzuje na diplomovú prácu [1] a rozširuje ju o praktické aplikácie, porovnania metód a ich vylepšenia.

V prvej kapitole práce si v skratke uvedieme teóriu k portfóliu, analýze portfólia a k stavovému modelovaniu.

V druhej kapitole práce popíšeme teóriu Kalmanovho filtra, postup pri stanovení počiatočných podmienok filtrácie a vysvetlíme ako Kalmanov filter použiť pre analýzu zloženia portfólia.

V tretej kapitole poukážeme na niektoré praktické aspekty použitia softwaru k dynamickej analýze. Ďalej sa zameriame na predpoklady Kalmanovho filtra a pozrieme sa na prípady, kedy predpoklady nebudú splnené. Okrem týchto prípadov si ukážeme aj praktické aplikácie filtra na rôznych druhoch umelých aj reálnych dát.

V štvrtej kapitole popíšeme teóriu k lokálne váženej regresii ako alternatívnej metóde dynamickej analýzy zloženia portfólia. Budeme hľadať vhodné nastavenie regresie a ukážeme si jej praktické aplikácie opäť na umelých aj reálnych dátach.

V záverečných kapitolách práce vysvetlíme pojem multikolinearita a vysvetlíme ako vplýva na výsledky filtrácie, zavedieme si porovnávacie kritéria pre porovnanie výsledkov oboch metód a pokúsime sa ponúknuť návod, na základe ktorého sa čitateľ bude pri vlastnej dynamickej analýze zloženia portfólia rozhodovať, ktorú metódu a s akými nastaveniami použije.

Na záver si ešte ukážeme spôsoby, ako v práci popísané metódy rozšíriť a vylepšiť, napríklad pomocou pridávania podmienok do filtrácie, alebo pomocou aplikácie štatistických testov detekcie zmien.

1. Základné informácie

Na úvod si vysvetlíme čo vlastne portfolio je, ako vzniká a prečo je potrebná jeho analýza. Popíšeme ako sa počítajú výnosy z akcií a ako z nich zložiť portfólio.

1.1 Portfólio

V ekonómii sa pojem portfólio používa pre súbor akcií, dlhopisov, podielových listov či cenných papierov jedného investora. Definície sa však v literatúre líšia a prispôsobujú. Obecné sa pri zostavovaní portfólia investor musí rozhodnúť, ktorý z parametrov - výnos, riziko, likvidita, je pre neho najdôležitejší. V skutočnosti ide o stanovenie pre investora prípustných hodnôt týchto parametrov, keďže v praxi sú výrazne korelované a prakticky nie je možné zvýšiť výnos portfólia bez toho, aby sa zvýšilo jeho riziko a podobne.

Matematicky pojem portfólio dobre vysvetľuje Modern Portfolio Theory (MPT) [8]. MPT je matematická formulácia konceptu diverzifikácie investícií, s cieľom vybrať súbor investičných aktív, ktorých spoločné riziko je nižšie ako pre každé aktívum samostatne. Toto je možné, keď sa hodnoty rôznych typov aktív menia rôznym spôsobom. Napríklad ceny na akciovom trhu sa pohybujú inak ako ceny na trhu s dlhopismi a súbor aktív týchto dvoch typov môže teoreticky znamenať nižšie riziko ako má každé aktívum zvlášť. Ale diverzifikácia znižuje riziko aj ak výnosy aktív nie sú negatívne korelované.

Matematicky MPT modeluje výnosy aktív pomocou normálneho rozdelenia (alebo všeobecnejšie ako elipticky rozdelené náhodné veličiny), riziko definuje ako štandardnú odchýlku výnosov a portfólio modeluje ako vážený súčet aktív. Výnos portfólia je potom vážený súčet výnosov jednotlivých aktív. V ďalších kapitolách využijeme tieto myšlienky ku konštrukcii (nie len) fiktívneho portfólia, ktoré následne budeme analyzovať.

1.2 Analýza portfólia

Existuje veľa analýz a metód, ktoré sa zaoberajú tým, ako správne zostaviť portfólio spĺňajúce základné pravidlá rozumného investovania (napr. Markowitzovo portfolio). My sa zameriame na analýzu už existujúceho portfólia. Postupne popíšeme a porovnáme metódy k určeniu zloženia portfólia z jeho výnosov a zameriame sa predovšetkým na tie, ktoré analyzujú portfólio v čase. Vstupné údaje budú pre nás výnosy jednotlivých zložiek, resp. výnosy zložiek, ktoré predpokladáme, že sú v portfóliu, keďže v realite často nemusíme poznať presné zloženie portfólia a celkové výnosy portfólia. Z výnosov zložiek budeme chcieť zistiť zloženie portfólia, t.j. váhy jednotlivých zložiek.

Na začiatok si vysvetlíme, čo sú vlastne výnosy. Označme cenu akcie (príp. hodnotu portfólia) v čase t ako S_t . Výnos akcie v čase t za čas Δ_t vypočítame z ceny akcie ako

$$vynos_t = \frac{S_t - S_{t-\Delta_t}}{S_{t-\Delta_t}}.$$

Na modelovanie výnosov sa často používajú logaritmické výnosy

$$\ln\left(\frac{S_t}{S_{t-\Delta t}}\right) = \ln\left(1 + \frac{S_t - S_{t-\Delta t}}{S_{t-\Delta t}}\right) \approx \frac{S_t - S_{t-\Delta t}}{S_{t-\Delta t}}.$$

Posledná aproximácia vyplýva z aproximácie

$$\ln(1+x) \approx x \text{ pre } x \approx 0.$$

Model pre výnosy akcií potom predpokladá, že jednotlivé výnosy sú nezávislé a sú z normálneho rozdelenia, čo využijeme v ďalších kapitolách pri modelovaní výnosov a zostavovaní umelého portfólia.

Z rovnice pre výnosy vidieť, že výnosy môžeme počítať za ľubovoľné časové obdobie. My budeme v práci využívať denné výnosy, t.j. výnosy medzi časmi t a $t-1$ a to aj v prípade reálnych cien akcií, ktoré sa udávajú len pre dni, kedy sa obchodovalo na burze (bez víkendov a štátnych sviatkov). To znamená, že výnos medzi piatkom a pondelkom nasledujúceho týždňa budeme považovať za jednodenný, napriek tomu, že reálne ubehli tri dni. Je to bežná konvencia používaná pri obchodovaní na burze.

Pozrime sa teraz na význam analýzy portfólia. V prípade, že sme vlastními, resp. obchodníkmi s daným portfóliom, asi nepotrebujeme analýzu na to, aby sme zistili zloženie. V prípade potenciálnych investorov je však zaujímavé pozrieť sa na to, ako sa menilo zloženie portfólia, do ktorého chceme investovať. Môže nás napríklad zaujímať, či výnos, ktorý vykazuje, nie je založený na priveľmi rizikovitom aktíve, alebo či si správca daného portfólia plní svoje povinnosti a riadi ho pre nás akceptovateľne.

Na modelovanie samotného portfólia z výnosov sa v praxi najčastejšie používa stavové modelovanie.

1.3 Stavové modelovanie

Stavové modelovanie slúži k popisu systému, ktorý nie je stacionárny v čase. Je obľúbené vďaka možnosti zahrnúť do systému chýbajúce pozorovania a tieto následne odhadovať. V práci budeme využívať dynamické lineárne modely, preto sa teraz zameriame na ich vysvetlenie.

Dynamické lineárne modely sa definujú pomocou dvoch rovníc. Rovnice pozorovaní a stavovej rovnici (v anglickej literatúre občas nazývaná aj evolučná rovnica). Stavová rovnica má tvar

$$x_t = F_t x_{t-1} + v_t,$$

kde x_t je n -dimenzionálny vektor, F_t je $n \times n$ dimenzionálna evolučná matica a v_t je náhodná veličina z normálneho rozdelenia. Rovnica pozorovaní vyzerá nasledovne

$$y_t = G_t x_t + w_t,$$

kde y_t je postupnosť pozorovaní, G_t je vektor alebo matica vysvetľujúcich premenných a w_t je opäť náhodná veličina z normálneho rozdelenia.

Dynamický lineárny systém je potom systém popísaný týmito dvoma rovnicami. Najjednoduchší model časových radov je model, kde G_t a F_t sú matice identít. My sa v práci budeme venovať o niečo zložitejšiemu modelu časových radov.

2. Kalmanov filter

V tejto kapitole si najskôr vysvetlíme čo je Kalmanov filter. Kalmanov filter je rekurzívny algoritmus, ktorý zo zašumených vstupných dát vyfiltruje štatisticky optimálny odhad stavového systému. Filter má množstvo aplikácií v rôznych oblastiach. Využíva sa napríklad v navigačných a riadiacich systémoch áut a lietadiel, v FM rádiách či iných komunikačných zariadeniach a v investovaní na analýzu portfólia. My sa zameriame na využitie pre analýzu portfólia. Na začiatok vysvetlíme príslušnú teóriu a zavedieme potrebné značenie. Následne ukážeme ako táto metóda funguje prakticky. Keďže v praktických aplikáciách využijeme software Mathematica, teoretická časť sa bude z časti opierať o dokumentáciu k tomuto softveru k príslušnej problematike [3] a z časti o [2].

2.1 Zavedenie Kalmanovho filtra

Kalmanov filter je technika, ktorá rekurzívne odhaduje nepozorované veličiny nazývané aj stavové premenné $\{x_t\}$ z pozorovaných časových radov $\{y_t\}$. Pozorovania zapíšeme nasledujúcou sústavou:

$$y_t = G_t x_t + w_t,$$

$$x_t = F_t x_{t-1} + v_t,$$

kde $\{y_t\}$ je pozorovaný časový rad a $\{x_t\}$ je stavová veličina. F a G sú známe matice alebo vektory, ktoré môžu byť závislé na čase. $\{w_t\}$ a $\{v_t\}$ predstavujú náhodné veličiny nezávislého (v_t nezávislé od v_s , pre každé t, s a v_t a w_t vzájomne nezávislé) gausovského bieleho šumu s nulovými strednými hodnotami a kovariančnými maticami $\mathbb{E}v_t v_s' = \delta_{ts} Q_t$, $\mathbb{E}w_t w_s' = \delta_{ts} R_t$.

Kalmanov filter udáva rekurzívne formule pre konštrukciu najlepšieho lineárneho odhadu \hat{x}_t^t , v zmysle strednej kvadratickej odchylky, stavu x_t a chybovej matice $P_t^t = \mathbb{E}(x_t - \hat{x}_t^t)(x_t - \hat{x}_t^t)'$ v aktuálnom čase t využívajúc všetky informácie dostupné do času t t.j. $Y^t = \{y_0, y_1, \dots, y_t\}$. Tieto formule majú tvar:

$$\hat{x}_t^t = \hat{x}_t^{t-1} + P_t^{t-1} G_t (G_t' P_t^{t-1} G_t' + R_t)^{-1} (y_t - G_t' \hat{x}_t^{t-1}),$$

$$P_t^t = P_t^{t-1} - P_t^{t-1} G_t (G_t' P_t^{t-1} G_t' + R_t)^{-1} G_t' P_t^{t-1},$$

kde

$$\hat{x}_t^{t-1} = F_t \hat{x}_{t-1}^{t-1},$$

$$P_t^{t-1} = F_t P_{t-1}^{t-1} F_t' + Q_t$$

sú predikované hodnoty skonštruované pre čas t v čase $t - 1$.

Štandardný Kalmanov filter predpokladá normálne rozdelenie reziduí w_t a v_t , t.j.

$$v_t \sim N(0, Q_t), w_t \sim N(0, R_t).$$

Potom \hat{x}_t^t je v skutočnosti najlepší odhad stavu x_t pretože platí

$$\hat{x}_t^t = \mathbb{E}(x_t|Y').$$

Túto skutočnosť nebudeme dokazovať, dôkaz je možné nájsť napríklad v [1]. Podmienka normálneho rozdelenia reziduí v praxi často nie je splnená preto sa Kalmanov filter „zrobustňuje“. Viac informácií k „zrobustňovaniu“ je možné nájsť v [2].

Všetky vyššie uvedené poznatky zhrnieme pre korektnosť do vety prevzatej z [1] a prispôbenej vyššie zavedenému značeniu. Vetu nebudeme dokazovať, dôkaz (presného pôvodného znenia) je možné nájsť v [4].

Veta 1. (*Kalmanov filter*)

Majme lineárny, konečne dimenzionálny, diskretný systém reprezentovaný rovnicami

$$y_t = G_t x_t + w_t,$$

$$x_t = F_t x_{t-1} + v_t,$$

definovanými pre $t \geq 0$. Náhodné procesy $\{v_t\}$ a $\{w_t\}$ sú nezávislé s normálnym rozdelením a platí pre ne:

$$\mathbb{E}(v_t) = 0 \quad \mathbb{E}(w_t) = 0 \quad \forall t \geq 0,$$

$$\mathbb{E}(w_t w_s') = \delta_{ts} R_t \quad \mathbb{E}(v_t v_s') = \delta_{ts} Q_t \quad \forall t, s \geq 0,$$

kde δ_{ts} predstavuje Kroneckerovo delta. Ďalej predpokladáme, že počiatočný stav x_0 je náhodná veličina s normálnym rozdelením a so strednou hodnotou \bar{x}_0 a variačnou maticou P_0 . Počiatočný stav x_0 je nezávislý na $\{v_t\}$ a $\{w_t\}$. Potom platí

$$\hat{x}_{t+1|t} = F_t \hat{x}_{t|t-1} + K_t (y_t - G_t \hat{x}_{t|t-1}),$$

s počiatočnou podmienkou

$$\hat{x}_{0|-1} = \bar{x}_0$$

a maticou K_t v tvare

$$K_t = F_t P_{t|t-1} G_t' (G_t P_{t|t-1} G_t' + R_t)^{-1},$$

kde predpokladáme, že inverzia matice $(G_t P_{t|t-1} G_t' + R_t)$ existuje. Ďalej potom

$$P_{t+1|t} = F_t [P_{t|t-1} - P_{t|t-1} G_t' (G_t P_{t|t-1} G_t' + R_t)^{-1} G_t P_{t|t-1}] F_t' + Q_t$$

s počiatočnou podmienkou

$$P_{0|-1} = P_0.$$

Pritom $\hat{x}_{t|t}$ a $P_{t|t}$ získame z rovníc

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + P_{t|t-1} G_t' (G_t P_{t|t-1} G_t' + R_t)^{-1} (y_t - G_t \hat{x}_{t|t-1}),$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} G_t' (G_t P_{t|t-1} G_t' + R_t)^{-1} G_t P_{t|t-1}.$$

Táto veta nám hovorí ako pomocou Kalmanových rekurzívnych vzorcov vypočítať odhad $\hat{x}_{t|t}$ a tiež príslušnú variačnú maticu $P_{t|t}$ za podmienky, že poznáme matice zo sústavy F_t , G_t , R_t , Q_t , počiatočné podmienky P_0 a \bar{x}_0 a pozorovania y do času t .

2.2 Počiatočné podmienky

K rekurzívnemu výpočtu Kalmanovho filtra sú potrebné počiatočné podmienky. V knihe [5] sú uvedené základné metódy pre určenie počiatočných podmienok a preto ich v tejto práci odvodzovať nebudeme, len ich v skrátenej forme popíšeme. Prvá metóda je veľmi reštriktívna, jej použitie je možné len za určitých podmienok. Tými sú stacionarita modelu v čase (teda F , G , R a Q sú nezávislé na t) a tiež stacionarita stavového vektora. Za týchto podmienok je možné odvodiť počiatočné hodnoty \bar{x}_0 a P_0 . Dostaneme

$$\mathbb{E}(x_0) = \bar{x}_0 = 0$$

a P_0 sa získa ako riešenie rovnice

$$P_0 = FP_0F' + Q.$$

V prípade, že stavová rovnica nie je stacionárna, tieto počiatočné podmienky nie je správne použiť. Častým riešením je preto voliť tzv. „diffuse priors“

$$\bar{x}_0 = 0,$$

$$P_0 = \kappa I,$$

kde $\kappa \rightarrow \infty$ a I je jednotková matica.

2.3 Kalmanov filter pre portfólio

Ukážeme si ešte, ako uvedená teória vyzerá v praxi, aké rozmery musia mať vstupné údaje Kalmanovho filtra a ako si filter jednoducho prispôbiť.

Začneme prevedením sústavy lineárneho dynamického systému do „reči portfólia“. Pripomeňme si ešte rovnice tohoto systému.

$$y_t = G_t x_t + w_t,$$

$$x_t = F_t x_{t-1} + v_t.$$

Ako sme už uviedli v predošlej kapitole, prvá rovnica je rovnica pozorovaní a druhá rovnica je stavová.

Predstavme si teraz, že máme portfólio zložené z n aktív (akcií). Potom y_t je vektor pozorovaní - výnosov celého portfólia, zatiaľ čo G_t je matica zložená z výnosov jednotlivých zložiek portfólia v čase a x_t sú váhy jednotlivých zložiek portfólia. F_t je jednotková (predpokladáme, že váha aktíva v čase t závisí len na váhe v čase $t - 1$ a prípadnom šume - inováciách).

Výššie uvedený popis už poskytuje lepšiu názornosť na praktickosť problematiky, musíme si ale ešte uvedomiť jednu zásadnú vec. Pre pevný čas t je x_t vektor váh jednotlivých zložiek portfólia. Model ale nedáva na hodnoty x_t žiadnu podmienku. V praxi požadujeme minimálne logickú podmienku na súčet váh 100% (v našom modeli 1, zložky x_t sú percentá udávané v desatinných číslach). Ďalšou podmienkou môže byť podmienka na nezápornosť váh. Záporné váhy predstavujú tzv. predaj nakrátko, ktorý pre zjednodušenie počítania nebudeme zakazovať. Pridáme len podmienku na súčet váh rovný 1. Pre pevný čas t a počet zložiek portfólia n má potom sústava tvar

$$\begin{pmatrix} y_t \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & g_{1t} & g_{2t} & \dots \\ 0 & 1 & 1 & \dots \end{pmatrix}}_{G_t} \begin{pmatrix} \alpha_t \\ x_{1t} \\ x_{2t} \\ \vdots \end{pmatrix} + \begin{pmatrix} w_{1t} \\ w_{2t} \\ \vdots \\ 0 \end{pmatrix}.$$

Táto metóda rozšírenia Kalmanovho filtra o podmienku na súčet váh sa v odbornej literatúre nazýva „Perfect Measurement“ [9]. Pre pevný čas t je y_t dvojzložkový vektor, kde druhá zložka je vždy 1 čím zabezpečíme, aby súčet váh zložiek portfólia bol rovný 1. G_t je matica typu $(2 \times (n + 1))$ kde n je počet zložiek portfólia, x_t a w_t sú $(n + 1)$ rozmerné vektory. α_t , prvá zložka vektoru x_t , sa nazýva Jensenova alpha. Jensenova alpha je ukazovateľ kvality riadenia portfólia. V prípade úspešne riadeného portfólia očakávame alpha kladné, v opačnom prípade záporné.

Takto vyzerá sústava pre jeden pevný čas t . My sa v práci budeme zaoberať dynamickým systémom a teda predpokladáme nekonštantnosť v čase t .

3. Praktické aplikácie Kalmanovho filtra

3.1 Výpočtové aspekty

Pred popisom samotnej aplikácie Kalmanovho filtra a výsledkov, ktoré sme filtráciou dosiahli si uvedieme niekoľko aspektov, ktoré výrazne ovplyvňujú presnosť požadovaných výstupov.

V celej výpočtovej časti tejto diplomovej práce využívame software Mathematica, verzia 8. Mathematica obsahuje balíček s názvom TimeSeries, ktorý obsahuje preddefinované funkcie k analýze časových radov, okrem iného aj funkcie KalmanFilter a LogLikelihood. Postup výpočtu začne využitím funkcie LogLikelihood, ktorá slúži k odhadu rozptylov inovácií metódou maximálnej vierohodnosti. Následne sa výstupy z tejto funkcie použijú ako vstupy pre funkciu KalmanFilter, ktorá slúži k filtrácii váh.

Pri aplikácii funkcie LogLikelihood, sme však narazili na niekoľko problémov. Prvým a najmenej závažným je konvencia, podľa ktorej má Mathematica kvalitné funkcie na hľadanie minima a nie maxima. Tento problém je len formálny a rieši sa jednoducho. Miesto hľadania maxima funkcie LogLikelihood hľadáme minimum funkcie mínus LogLikelihood.

Druhý problém je už o niečo závažnejší. Minimalizačná funkcia FindMinimum (defaultná minimalizačná funkcia v dokumentácii) sa ukázala ako veľmi citlivá na zadané vstupy. Vstupom funkcie sú voliteľné, my sme zadali interval, v ktorom sa má pohybovať s hodnotami premennej pri hľadaní minima. Interval sme nastavili nasledovne - vieme, že rozptyl je vždy nezáporné číslo, spodná hranica intervalu je preto 0, hornú hranicu sme nastavili na 1, aby filtráciu nijak neobmedzovala (hodnoty rozptylov inovácií sa blížia skôr k 0 ako k 1). Ďalším voliteľným vstupom funkcie je štartovacia hodnota. Táto sa ukázala ako rozhodujúca v aplikácii. Funkcia FindMinimum využíva gradientnú metódu minimalizácie a preto v závislosti na štartovacej hodnote dospeje k nejakému minimu, často lokálnemu, na hranici prehľadávaného priestoru a keďže už sa nemá kam pohnúť, prehlási toto za minimum. Je preto dôležité zvoliť správnu štartovaciu hodnotu v zmysle správneho rádu (t.j. hľadáme n pre štartovaciu hodnotu typu 10^{-n}).

V prípade, že my sami generujeme dáta, poznáme vstupné parametre a vieme akú štartovaciu hodnotu zadať, aby sme dotali čo najpresnejšie výsledky. V opačnom prípade využijeme jednoduchý algoritmus (*For-cyklus* v Mathematice) na nájdenie požadovanej hodnoty, ktorý vyskúša všetky rády od 0.1 po 10^{-10} :

Pre všetky x_0 od 0.1 po 10^{-10} :

- Aplikuj FindMinimum so štartovacou hodnotou x_0
- Ulož minimum a štartovaciu hodnotu

Z uložených hodnôt minima potom vyberieme to najmenšie a jemu prislúchajúcu štartovaciu hodnotu použijeme k filtrácii.

Ďalším problémom je odhad rozptylov inovácií na 0. Vyššie sme uviedli, že funkcia FindMinimum hľadá minimum pre rozptyly inovácií v nami zadanom intervale $(0, 1)$. Problém nastane, ak sú tieto odhady práve na hranici intervalu 0. Minimum síce nájdeme, ale v ďalšom kroku funkcia KalmanFilter nespočíta hodnoty filtrácie, pretože potrebuje invertovať maticu, ktorá kvôli nulovým hodnotám rozptylov inovácií invertovať nejde. Problém sme vyriešili zväčšením spodnej hranice intervalu na 10^{-10} . Čím došlo k inej chybe - Mathematica defaultne počíta s presnosťou na 16 desatinných miest, my sme jej ale zadali len 10 desatinných miest. Prenastavením spodnej hranice na 10^{-17} sme chybovú hlášku odstránili a presnosť výsledkov sa tým zlepšila. Vyskúšali sme preto ešte zväčšiť parameter WorkingPrecision (presnosť výpočtu) a s ním (adekvátne) aj spodnú hranicu intervalu postupne na 20, 30, 40, 50, 60 (desatinných miest). S prechodom na WorkingPrecision rovné 50 sa ešte výsledky zlepšovali (a výpočtová doba predlžovala), s ďalej rastúcou presnosťou sme už zlepšenie nedosiahli.

Výsledkom týchto cvičení je vhodne nastavená funkcie FindMinimum.

3.2 Ideálny prípad

Na začiatok si ukážeme, ako funguje Kalmanov filter v ideálnom prípade kedy máme splnené všetky predpoklady. Predpoklady modelu zahŕňajú: normálne rozdelenie šumov v_t a w_t a nezávislosť v_t na w_t . Ďalšie predpoklady hovoria o vlastnostiach počiatočného stavu. Tieto budeme v celej práci považovať za splnené.

Využijeme poznatky z prvej kapitoly o modelovaní výnosov a zostavíme si vlastné, umelé portfólio presne podľa predpokladov Kalmanovho filtra. Najskôr vygenerujeme z náhodného rozdelenia $N(0, 0.05^2)$ tri sady výnosov dĺžky 504 (táto dĺžka odpovedá dĺžke reálnych dát, ktoré budeme neskôr používať, preto ju používame už tu, pre možnosť porovnávania). Následne si zvolíme váhy zložiek (volíme ako 0.4+náhodná prechádzka vygenerovaná z $N(0, 0.005^2)$, 0.1+náhodná prechádzka vygenerovaná z $N(0, 0.005^2)$ a 1 mínus prvé dve váhy. V príklade máme totiž zavedenú podmienku na súčet váh portfólia rovný jednej. K takto zloženému portfóliu pridáme pozorovací šum, opäť z rozdelenia $N(0, 0.005^2)$.

Týmto sme zostavili jednoduché portfólio, ktoré spĺňa všetky predpoklady Kalmanovho filtra. Pre prehľadnosť a jednoduchú orientáciu v práci ešte zhrnieme použité parametre príkladu do tabuľky 3.1. Takúto tabuľku nájde čitateľ pri každom príklade, aby nemusel vyhľadávať v texte, aké parametre sme použili.

počet zložiek portfólia	3
rozdelenie náh. prechádzky	$N(0, 0.005^2)$
rozdelenie pozorovacieho šumu	$N(0, 0.005^2)$
Jensenovo alpha zahrnuté	áno
výnosy portfólia	generované z $N(0, 0.05^2)$

Tabuľka 3.1: Parametre príkladu 1.

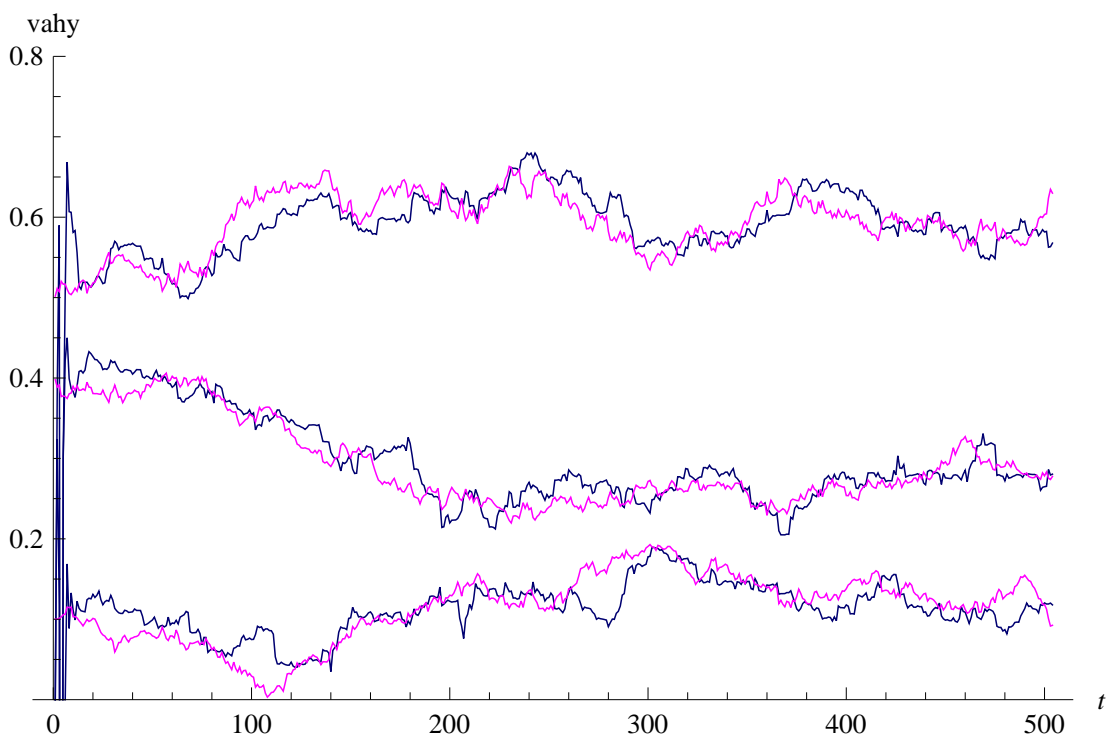
Upresnime ešte, že Jensenovo aplha je v celej tejto kapitole konštantne rovné 0 a že v celej práci používame rozšírenie príkladu, kedy požadujeme súčet váh rovný jednej.

Pozrieme sa ako vyzerajú výsledky prvého príkladu. Kalmanov filter najskôr odhadol rozptyly náhodných prechádzok a pozorovacieho šumu a to nasledovne:

σ_α^2	0
$\sigma_{x_1}^2$	$2.51714 \cdot 10^{-5}$
$\sigma_{x_2}^2$	$2.50598 \cdot 10^{-5}$
$\sigma_{x_3}^2$	$2.54223 \cdot 10^{-5}$
σ_w^2	$1.56042 \cdot 10^{-5}$

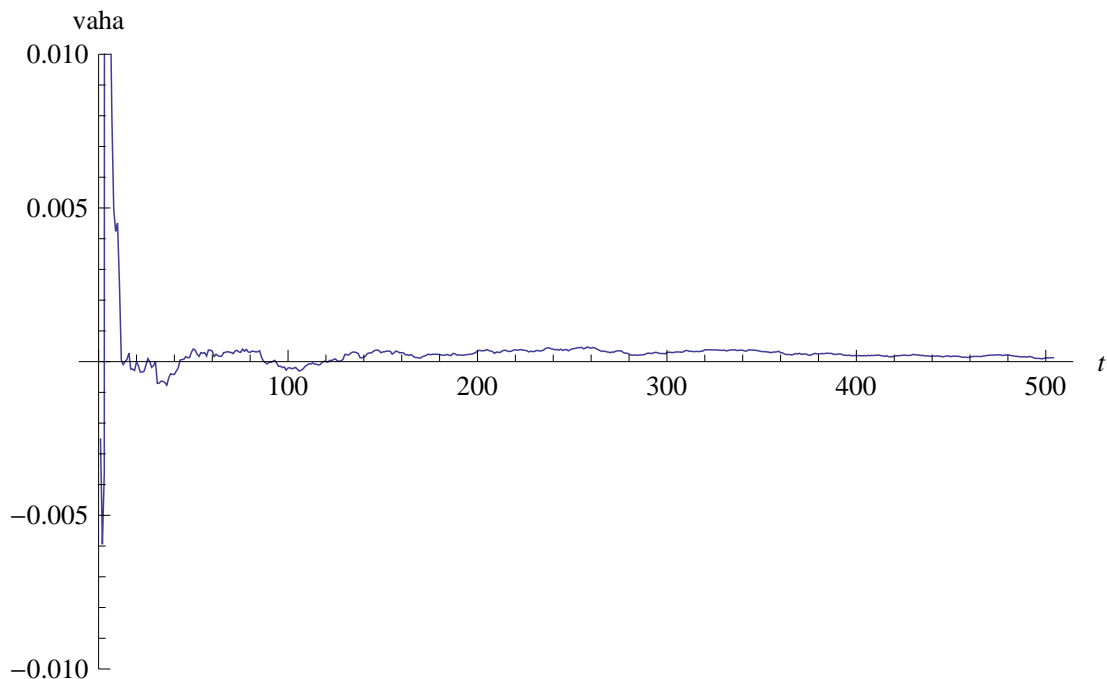
Tabuľka 3.2: Odhady rozptylov v ideálnom prípade.

Skutočný rozptyl nami vygenerovanej náhodnej prechádzky t.j. rozptyl $\sigma_{x_1}^2$ bol $2.5 \cdot 10^{-5}$. Vidíme, že filter tento rozptyl pre všetky inovácie zložiek portfólia správne odhadol, rozptyl pozorovacieho šumu zmenšil.



Obr. 3.1: Graf vyfiltrovaných (modré) a skutočných (ružové) váh v ideálnom prípade.

Na obrázku 3.1 vidíme vyfiltrované odhady váh (modré čiary) priamo proti skutočným váham (ružové čiary). Vidíme, že vyfiltrované odhady sú blízko skutočným váham. Pozrime sa ešte na graf Jensenovho alpha.



Obr. 3.2: Graf Jensenovho alpha pre ideálny prípad.

Kedže sme pre Jensenovo alpha žiadne vstupné údaje modelu neposkytli, predpokladáme, že vyfiltrované aplha bude nulové rovnako ako aj jeho rozptyl.

Rozptyl filter odhadol nulový a podľa x-ovej osi grafu 3.2 vidíme, že aj hodnoty Jensenovho alpha sa pohybujú blízko nuly.

Skôr ako spustíme ďalšiu filtráciu, zavedieme si ešte kritériá pre porovnávanie výsledkov, aby sme nemuseli usudzovať o presnosti filtrácie z grafov. Ako prvé kritérium si určíme súčet štvorcových odchýliek

$$\sum_{i=1}^m \sum_{j=1}^n (\hat{x}_{ij} - x_{ij}),$$

kde m je počet zložiek portfólia a n je dĺžka časového radu. Druhé kritérium, ktoré použijeme bude počet nesprávnych odhadov. Nesprávny odhad v tomto kontexte znamená, že vyfiltrovaná váha sa od skutočnej líši o viac ako 0.01 a to v absolútnej hodnote, t.j.

$$\sum_{i=1}^m \sum_{j=1}^n I(|\hat{x}_{ij} - x_{ij}| < 0.01).$$

Obe kritériá uvedieme vždy absolútne, ale pre možnosť porovnávania rôznych príkladov aj relatívne, vzhľadom k celkovému počtu vyfiltrovaných hodnôt.

Pre ideálny prípad vychádza súčet štvorcov 88.1652. Vychádzajúc z grafických výsledkov na obrázku 3.1, takto vysoké číslo môže byť výsledkom vplyvu prvých pár hodnôt. Túto domnienku potvrdíme napočítaním súčtu štvorcov s vynechaním prvých 10 pozorovaní. Modifikovaním definície súčtu štvorcov o vynechanie prvých 10 pozorovaní dostaneme hodnotu 0.969706. Odstránením pozorovaní, ktoré síce mali na kritérium najväčší vplyv, avšak skôr deformujúci, dostávame novú definíciu súčtu štvorcov, ktorá nám poskytne presnejšie výsledky v zmysle možnosti porovnania vývoja časových radov. Relatívny počet štvorcov napočítame vydelením absolútneho počtom zahrnutých pozorovaní, v tomto prípade 1482 (3 časové rady dĺžky 494).

Pre ideálny prípad so zahrnutým Jensenovým alpha nadobúdajú kritériá hodnoty uvedené v tabuľke 3.3. Relatívnu hodnotu pre počet nesprávnych odhadov sme získali vydelením absolútneho počtu číslom 1512 (3 časové rady dĺžky 494).

súčet štvorcov	0.969706
relatívny súčet štvorcov	0.000654323
počet nesprávnych odhadov	1036
relatívny počet nesprávnych odhadov	0.685185

Tabuľka 3.3: Hodnoty kritérií v ideálnom prípade s Jensenovým alpha.

3.3 Nesplnenie predpokladov filtra

V prvom príklade sme si ukázali, ako funguje Kalmanov filter v dokonalom prípade, keď sú splnené všetky jeho predpoklady. V reálnych príkladoch však často dochádza k výraznému porušeniu predpokladov (väčšinou hneď viacerých). Preto si teraz ukážeme, ako sa Kalmanov filter správa, ak sa menia podmienky a predpoklady prestávajú byť splnené. Pokúsime sa zistiť závažnosť týchto zmien a tiež to, ktorá konkrétna zmena spôsobuje filtru najväčšie problémy. V tejto podkapitole sa budeme venovať postupnému analyzovaniu Kalmanovho filtra pre prípady, kedy dôjde k porušeniu niektorého predpokladu. Filter požaduje, aby náhodné prechádzky inovácií a pozorovací šum boli normálne rozdelené a nezávislé, t.j. v_t je nezávislé na w_t .

3.3.1 Závislosť náhodnej prechádzky inovácií a pozorovacieho šumu

Začnime predpokladom nezávislosti. Čo sa stane, ak by pozorovací šum závisel na náhodných prechádzkach inovácií?

Skôr ako sa pustíme do modelovania a analyzovania ešte jedna poznámka. Pre účely tejto kapitoly budeme používať umelé dáta, ktoré si vygenerujeme tak, ako budeme potrebovať. Odôvodnenie je jednoduché - budeme mať istotu, že vieme ako dáta vyzerajú a aké majú vlastnosti.

Vráťme sa k otázke, ako reaguje Kalmanov filter, ak má porušený predpoklad na nezávislosť. Vygenerujeme tri časové rady výnosov z normálneho rozdelenia $N(0, 0.05^2)$. Následne si zvolíme váhy zložiek (volíme ako 0.4+náhodná prechádzka vygenerovaná z $N(0, 0.005^2)$, označme ju a , 0.1+náhodná prechádzka vygenerovaná z $N(0, 0.005^2)$, označme ju b a 1 mínus prvé dve váhy. V príklade máme zavedenú podmienku na súčet váh portfólia rovný jednej. K takto zloženému portfóliu pridáme pozorovací šum v tvare $\frac{\sqrt{2}}{2} \cdot a + \frac{\sqrt{2}}{2} \cdot b$ - kombinácia inovácií váh portfólia tak, aby zostali parametre pozorovacieho šumu nezmenené oproti ostatným príkladom (z vety o súčte normálnych rozdelení, napr. veta 3.16 (i) na str. 57 v [12]). Parametre tohoto príkladu uvádzame v tabuľke 3.4.

počet zložiek portfólia	3
rozdelenie náh. prechádzky	$N(0, 0.005^2)$
rozdelenie pozorovacieho šumu	$N(0, 0.005^2)$, závislé na náh. prechádzkach inovácií
Jensenovo alpha zahrnuté	áno
výnosy portfólia	generované z $N(0, 0.05^2)$

Tabuľka 3.4: Parametre príkladu 2 pre w_t závislé na v_t .

Pozrieme sa, ako vyzerajú výsledky v tomto prípade. Kalmanov filter najskôr odhadol rozptyly náhodných prechádzok a pozorovacieho šumu, a to nasledovne:

σ_α^2	0
$\sigma_{x_1}^2$	$2.48401 \cdot 10^{-5}$
$\sigma_{x_2}^2$	$2.43617 \cdot 10^{-5}$
$\sigma_{x_3}^2$	$2.51365 \cdot 10^{-5}$
σ_w^2	$1.38146 \cdot 10^{-5}$

Tabuľka 3.5: Odhady rozptylov pre w_t závislé na v_t .

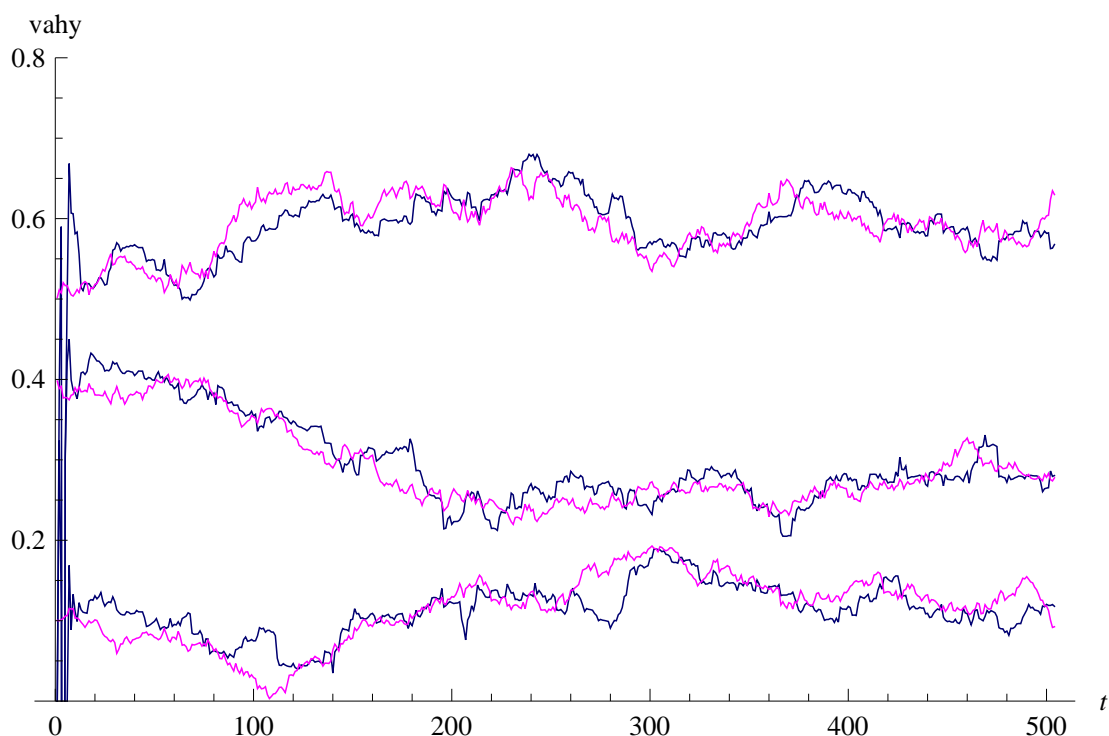
Z grafického výsledku na obrázku 3.3 je vidieť, že závislý šum nijako neovplyvnil výsledok filtrácie. Obrázok porovnáva výsledky filtrácie (modré čiary) so skutočnými váhami (ružové čiary).

Hodnoty kritérií sú v tomto prípade uvedené v tabuľke 3.6.

súčet štvorcov	0.892492
relatívny súčet štvorcov	0.000602221
počet nesprávnych odhadov	1035
relatívny počet nesprávnych odhadov	0.684524

Tabuľka 3.6: Hodnoty kritérií pre w_t závislé na v_t .

Pripomeňme, že súčet štvorcov počítame bez prvých 10 pozorovaní (ak by sme ich zahrnuli, súčet štvorcov by vyšiel 32.8301). Výsledky filtrácie sú porovnateľné s ideálnym prípadom. Porušenie predpokladu nezávislosti pozorovacieho šumu a inovácií váž filtráciu výrazne nezhorší.



Obr. 3.3: Graf vyfiltrovaných a skutočných váh pre w_t závislé na v_t .

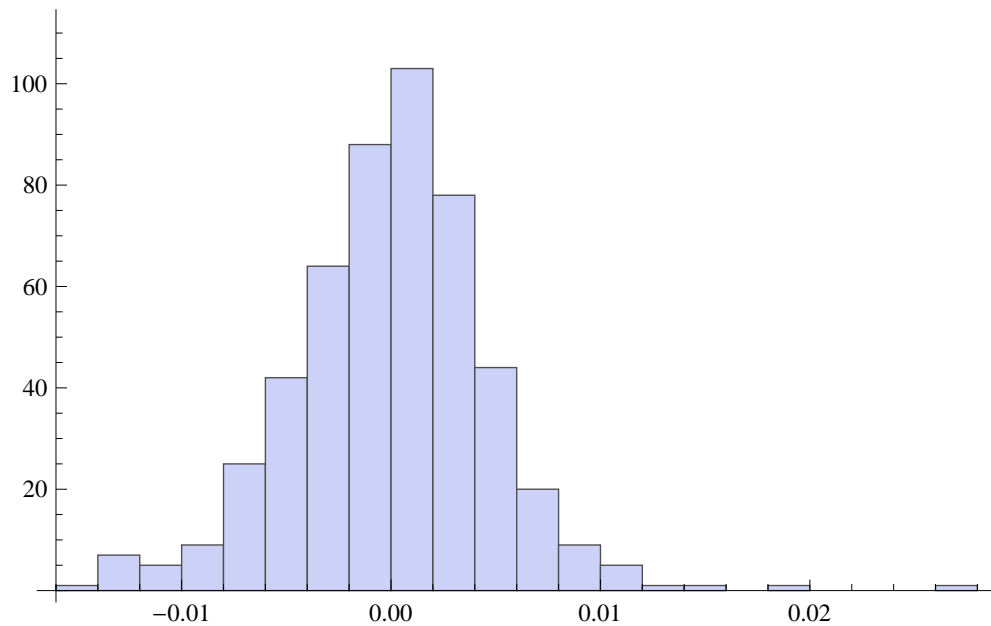
3.3.2 Porušenie normality šumov

Zamerajme sa teraz na druhý predpoklad, ktorým je normalita rozdelenia inovácií. Čo sa stane, ak inovácie váh nebudú z normálneho rozdelenia? Čo ak by boli napríklad zo Studentovho rozdelenia?

Pri voľbe parametrov Studentovho rozdelenia musíme mať stále na pamäti, že simulujeme inovácie váh. Nechceme preto príliš veľké hodnoty a hodnoty nad 1 už vôbec nie. Volíme preto rozdelenie $StudentT(0, 0.0041, 6)$. Stredná hodnota tohto rozdelenia je 0, škálovací parameter 0.0041 udáva veľkosti vygenerovaných hodnôt a počet stupňov voľnosti 6 udáva, ako veľmi chceme ťažké chvosty. Voľba týchto parametrov nám zabezpečuje, že rozdelenie má ťažšie chvosty ako normálne rozdelenie používané doteraz, zároveň ale stále spĺňa logickú predstavu o inovácii váh. Zároveň má toto rozdelenie porovnateľné parametre (strednú hodnotu a smerodiatnú odchýlku) s normálnymi rozdeleniami použitými v predošlých príkladoch. Stredná hodnota tohto rozdelenia je nulová ako aj predtým, smerodiatná odchýlka je 0.00502145. Histogram takto vygenerovaných inovácií je na obrázku 3.4, koeficient špicatosti tohto rozdelenia je 5.56266.

V tabuľke 3.7 uvádzame parametre tohto príkladu. Pozrime sa, ako vyzerajú odhady rozptylov inovácií v tomto prípade. Tieto môžeme vidieť v tabuľke 3.8.

Odhady rozptylov zodpovedajú očakávaniam. Na obrázku 3.5 vidieť, že ani s takouto zmenou nemá Kalmanov filter problém. Vyfiltrované váhy (modré čiary) zodpovedajú tým skutočným (ružové čiary). Tento príklad sme uviedli len pre ilustráciu toho, že podmienka normálneho rozdelenia šumov môže byť porušená, a filter napriek tomu funguje.



Obr. 3.4: Histogram inovácií zo Studentovho rozdelenia.

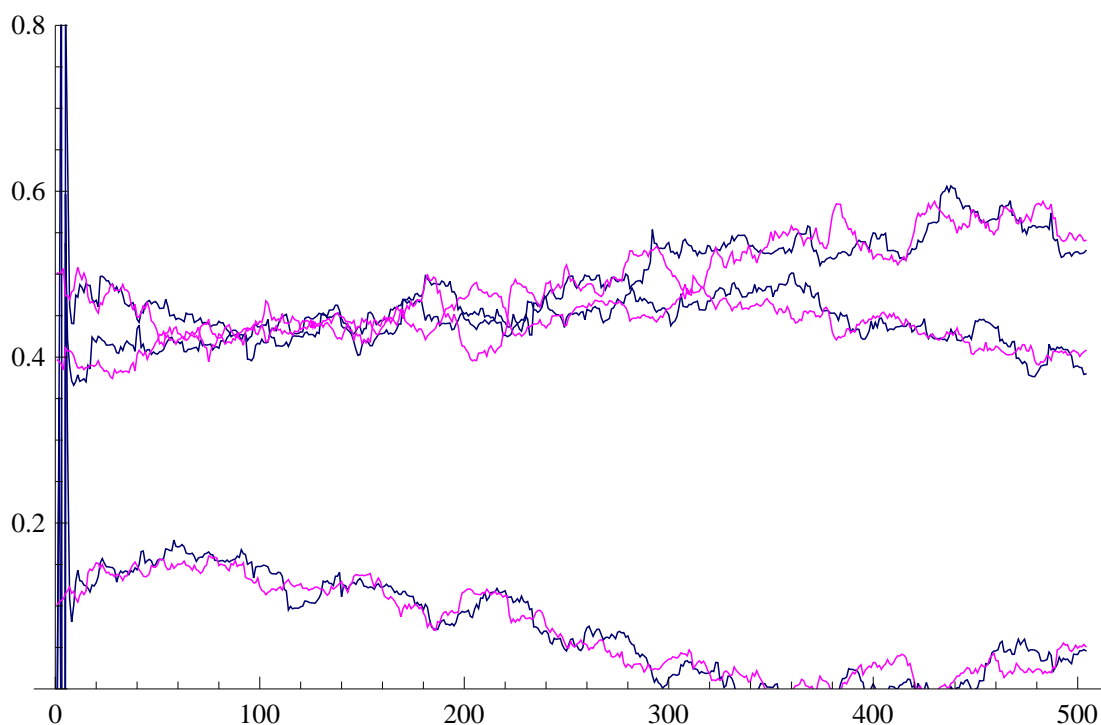
počet zložiek portfólia	3
rozdelenie náh. prechádzky	$StudentT(0, 0.0041, 6)$
rozdelenie pozorovacieho šumu	$StudentT(0, 0.0041, 6)$
Jensenovo alpha zahrnuté	áno
výnosy portfólia	generované z $N(0, 0.05^2)$

Tabuľka 3.7: Parametre príkladu 3 pre inovácie zo Studentovho rozdelenia.

σ_α^2	0
$\sigma_{x_1}^2$	$2.41636 \cdot 10^{-5}$
$\sigma_{x_2}^2$	$2.47676 \cdot 10^{-5}$
$\sigma_{x_3}^2$	$2.46813 \cdot 10^{-5}$
σ_w^2	$1.53057 \cdot 10^{-5}$

Tabuľka 3.8: Odhady rozptylov pre inovácie zo Studentovho rozdelenia.

Hodnoty kritérií sú v tomto prípade opäť porovnateľné s ideálnym prípadom. Jedna odchylka tu ale je. Je ňou hodnota súčtu štvorcov bez vynechania prvých 10 pozorovaní. Súčet štvorcov v tomto prípade vychádza 178.45, čo je takmer 2 krát viac ako pre ideálny prípad. Porušenie normality rozdelenia inovácií preto zrejme zhorší úvodnú prácu filtra a sťaží mu podmienky. Napriek tomu ostatné výsledky už sú kvalitné a porovnateľné s filtrom pri splnených predpokladoch.



Obr. 3.5: Graf vyfiltrovaných (modré čiary) a skutočných váh (ružové čiary) pre šum zo Studentovho rozdelenia.

súčet štvorcov	0.635765
relatívny súčet štvorcov	0.000428991
počet nesprávnych odhadov	1009
relatívny počet nesprávnych odhadov	0.667328

Tabuľka 3.9: Hodnoty kritérií pre inovácie zo Studentovho rozdelenia.

3.3.3 Šumy so skokmi

V reálnych situáciách sa však často stretávame ešte aj s iným typom šumu - so skokmi. Je to prípad, kedy dôjde k výraznejšej zmene portfólia, napríklad kompletnému odpredaniu 20% akcií jednej spoločnosti a pod. Keďže takéto zmeny nie sú nezvyčajné, namodelujeme si aj túto situáciu.

Tu sa trochu viac pohráme s modelovaním váh. Najskôr, tak ako doteraz, vygenerujeme šum z normálneho rozdelenia $N(0, 0.005^2)$. K vygenerovanému šumu budeme chcieť v náhodných časoch pridať skoky. To spravíme nasledovne. Vygenerujeme 504 zložiek z rovnomerného rozdelenia ($U(-0.1, 0.1)$) a 504 z Bernoulliho rozdelenia ($Alt(0.01)$). Vektor rovnomerného a alternatívneho rozdelenia prenasobíme po zložkách, čím získame hodnoty skokov (na miestach kde vo vektore alternatívneho rozdelenia boli jednotky). Tento vektor pripočítame po zložkách k vektoru pôvodného šumu (z normálneho rozdelenia). Dostali sme šum so skokmi, ktorý teraz pripočítame k zvolenej váhe. Volíme portfólio rovnako ako v predošlých prípadoch. Jedna váha bude na začiatok 0.3 a v každom kroku k nej pripočítame jednu zložku vektora šumu. Druhá zložka portfólia má váhu 0.1 plus šum a tretia má potom váhu 1 mínus váha prvých dvoch zložiek. Parametre tohto

príkladu sú zhrnuté v tabuľke 3.10.

počet zložiek portfólia	3
rozdelenie náh. prechádzky	$N(0, 0.005^2) + \text{skoky}$
rozdelenie pozorovacieho šumu	$N(0, 0.005^2)$
Jensenovo alpha zahrnuté	áno
výnosy portfólia	generované z $N(0, 0.05^2)$

Tabuľka 3.10: Parametre príkladu 4 pre inovácie so skokmi.

Výberové rozptyly takto nagenerovaných inovácií so skokmi sú $5.01599 \cdot 10^{-5}$ pre váhu v_1 a $9.04591 \cdot 10^{-5}$ pre váhu v_2 . K nim prislúchajúce, filtrom odhadnuté, rozptyly uvádzame v tabuľke 3.11. Tieto sú väčšie ako skutočné rozptyly čo je dôsledkom vysporiadania sa filtra so skokmi.

σ_α^2	0
$\sigma_{x_1}^2$	$1.68 \cdot 10^{-5}$
$\sigma_{x_2}^2$	$1.88 \cdot 10^{-5}$
$\sigma_{x_3}^2$	$2.00 \cdot 10^{-5}$
σ_w^2	$4.46 \cdot 10^{-5}$

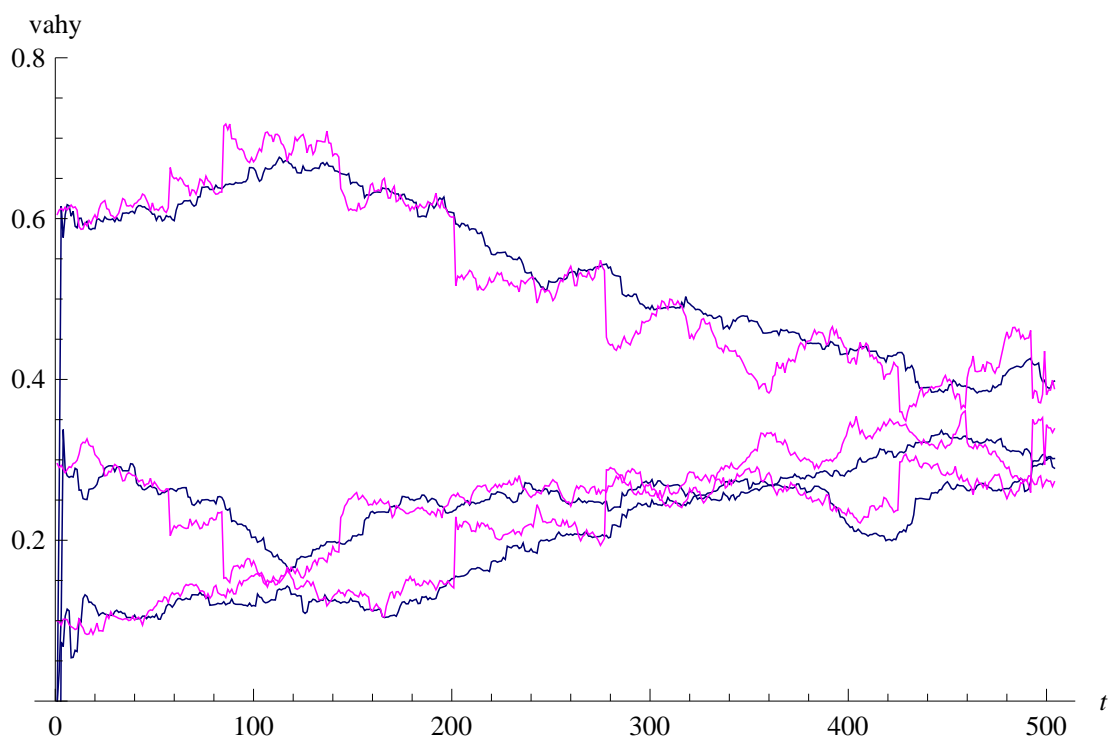
Tabuľka 3.11: Odhady rozptylov pre inovácie so skokmi.

Výsledok filtrácie uvádzame na obrázku 3.6, kde ružové čiary sú skutočné váhy, modré čiary sú ich vyfiltrované odhady. Z obrázku vidieť, že počas filtrácie dochádza k pomalej konvergencii k správnym hodnotám po okamžiku skoku. Táto skutočnosť sa odráža aj na výsledkoch hodnotenia pomocou kritérií v tabuľke 3.12.

súčet štvorcov	1.27796
relatívny súčet štvorcov	0.000862323
počet nesprávnych odhadov	1056
relatívny počet nesprávnych odhadov	0.698413

Tabuľka 3.12: Hodnoty kritérií pre inovácie so skokmi.

Záverom z týchto pozorovaní je skutočnosť, že aj keď nie sú splnené predpoklady Kalmanovho filtra, výsledky filtrácie sú stále uspokojivé. Výraznejšie



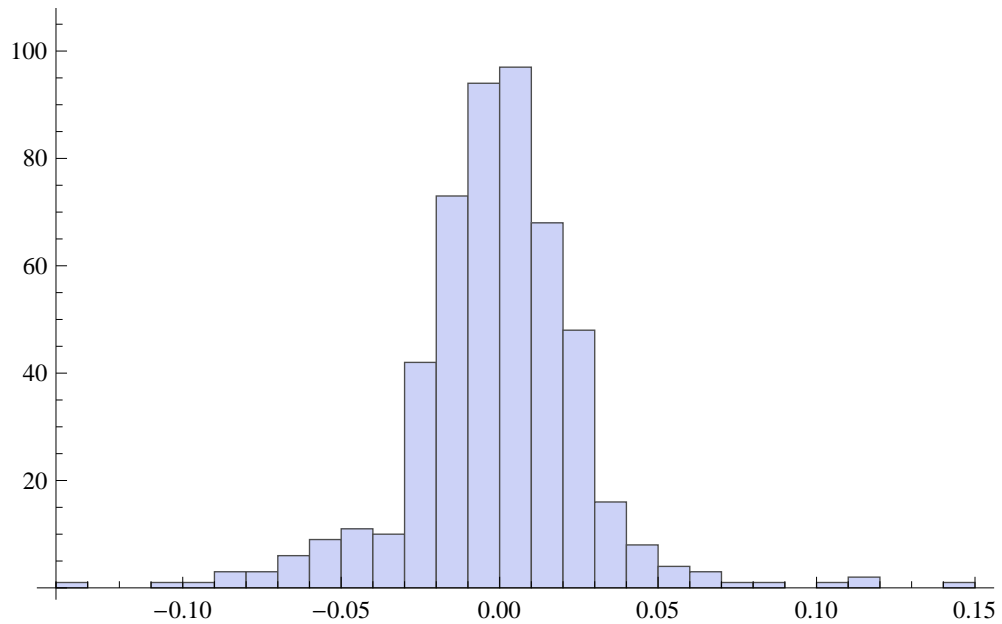
Obr. 3.6: Graf skutočných (ružové) a vyfiltrovaných váh (modré) pre inovácie s skokmi.

nepresnosti odhadovania váh sú spôsobené skokmi v inováciách, čo môže viesť k nepresným výsledkom v prípade, že skokov bude priveľa a filtru nebude medzi nimi poskytnutý dostatočný počet pozorovaní na dokonvergovanie k správnej váhe.

3.4 Príklad portfólia s reálnymi výnosmi

V tejto podkapitole si ukážeme, ako sa Kalmanov filter správa, ak sa menia podmienky (pridávajú sa zložky filtra) a postupne sa mu zadávajú reálne dáta. Pozrieme sa tiež, čo vplýva na kvalitu vyfiltrovaných odhadov.

Na začiatok zostavíme opäť jednoduchý príklad, tentokrát ale už s reálnym portfóliom. Z internetovej stránky pražskej burzy sme si stiahli ceny akcií za posledné dva roky (2010 a 2011) a napočítali sme si z nich výnosy. V tomto príklade použijeme výnosy Erste Group, VIG (Vienna Insurance Group) a Komerční banky (výber práve týchto akcií je čiste praktický, ich výnosy sme stiahli kvôli reálnemu príkladu, ku ktorému sa dostaneme neskôr v tejto kapitole). Tieto výnosy už nemajú ideálne normálne rozdelenie, ich rozdelenie je špicatejšie (viď histogram pre Erste Group na obrázku 3.7), s koeficientom špicatosti 7.5. Z týchto výnosov umelo zložíme portfólio tak, ako v ideálnom prípade na začiatku kapitoly. Časové rady pre reálne výnosy sú dĺžky 504, čo zodpovedá počtu pracovných dní za dva roky, za ktoré sme výnosy stiahli. V tabuľke 3.13 zhrnieme parametre príkladu.



Obr. 3.7: Histogram výnosov Erste Group.

počet zložiek portfólia	3
rozdelenie náh. prechádzky	$N(0, 0.005^2)$
rozdelenie pozorovacieho šumu	$N(0, 0.005^2)$
Jensenovo alpha zahrnuté	áno
výnosy portfólia	reálne

Tabuľka 3.13: Parametre príkladu 5 pre portfólio s 3 reálnymi zložkami.

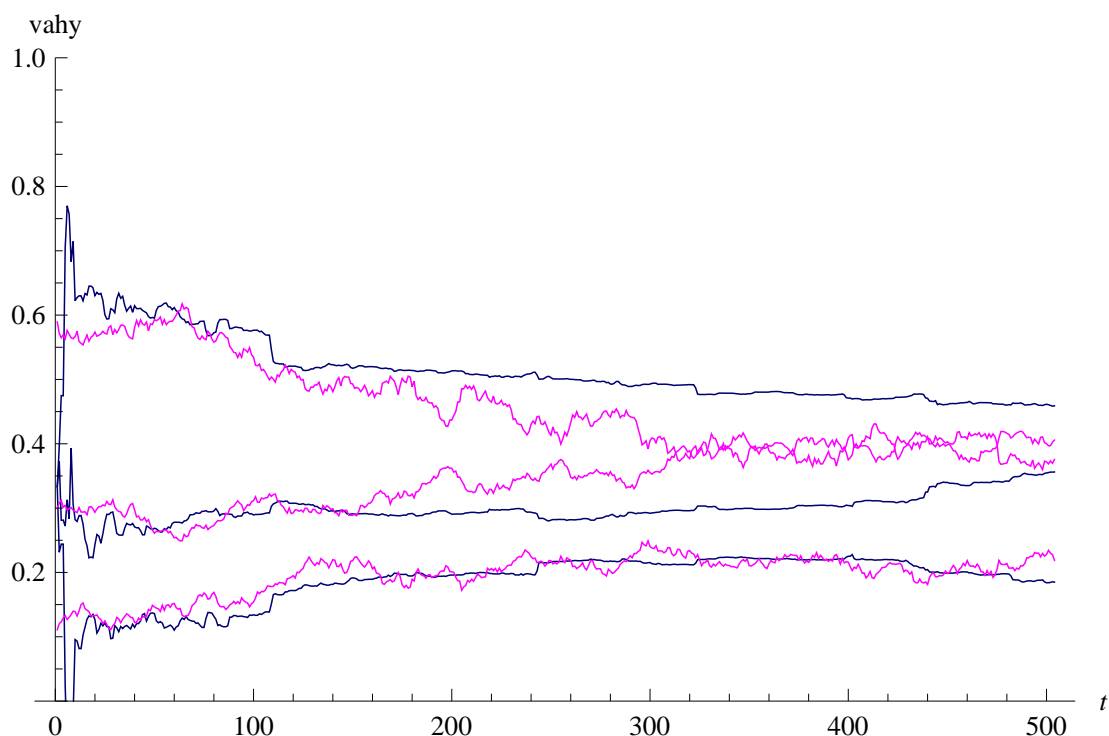
Opäť sa pozrieme na všetky výsledky poskytnuté Kalmanovým filtrom - odhady rozptylov šumu, vyfiltrované váhy a Jensenovo alpha. Odhady rozptylov sú zobrazené v tabuľke 3.14.

σ_α^2	$1.20000 \cdot 10^{-5}$
$\sigma_{x_1}^2$	$0.025555 \cdot 10^{-5}$
$\sigma_{x_2}^2$	$0.04595 \cdot 10^{-5}$
$\sigma_{x_3}^2$	$0.03673 \cdot 10^{-5}$
σ_w^2	$1.00000 \cdot 10^{-17}$

Tabuľka 3.14: Odhady rozptylov pre portfólio s 3 reálnymi zložkami.

Už z tabuľky odhadov rozptylov inovácií vidieť, že výsledky filtrácie nie sú presné. Filter zväčšil rozptyl inovácie Jensenovho alpha a zmenšil ostatné rozptyly. Grafické výsledky sú na obrázku 3.8.

Pozrime sa ešte na hodnoty Jensenovho alpha. Vieme, že toto by malo byť konštantne nulové, filter ale odhadol rozptyl jeho inovácie nenulový.



Obr. 3.8: Graf skutočných (ružové) a vyfiltrovaných váh pre portfólio s 3 reálnymi zložkami.

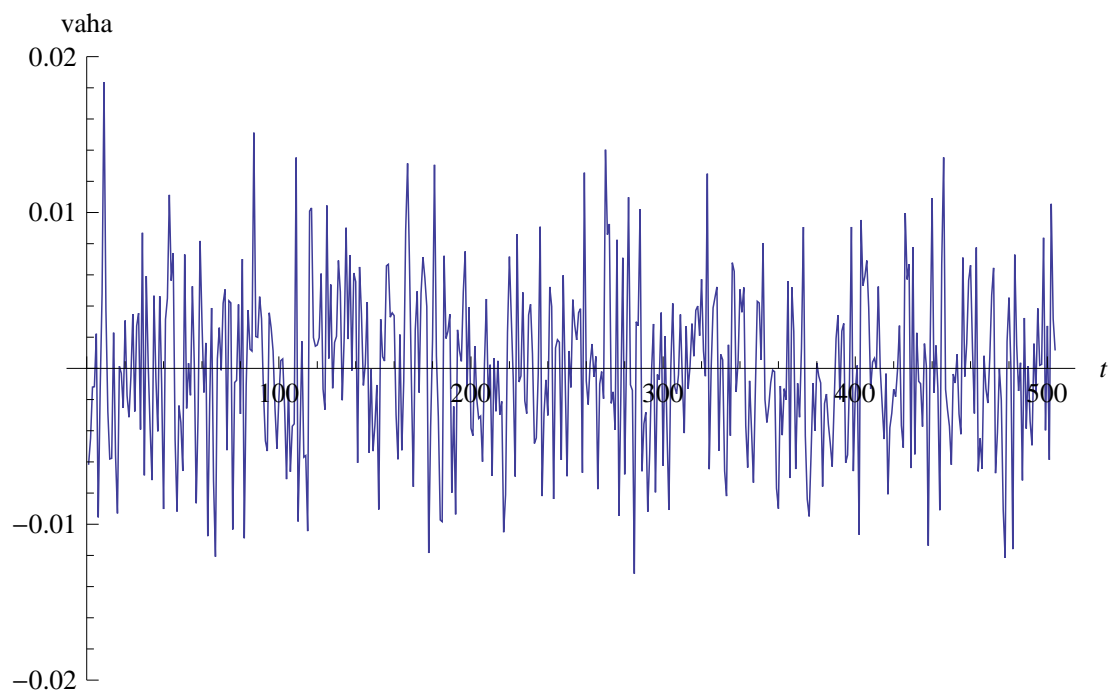
To viedlo k tomu, že zmeny, ktoré sa v skutočnosti diali vo váhach zložiek portfólia sa odrazili v inováciách Jensenovho alpha. V tomto prípade Jensenovo alpha (negatívne) ovplyvňuje konvergenciu filtrácie.

Tento jav je zrejme spôsobený výrazne nenulovými koreláciami medzi cenami jednotlivých akcií (napriek tomu, že predpoklady filtra o korelácii nehovoria). Všetky tri zvolené spoločnosti sú veľké a významné skupiny na trhu a reakcie ich cien na zmeny trhu nie sú nezávislé. Pozrieme sa na korelačnú maticu cien týchto akcií.

$$\begin{pmatrix} 1 & 0.897816 & 0.850344 \\ 0.897816 & 1 & 0.860883 \\ 0.850344 & 0.860883 & 1 \end{pmatrix}$$

Ako vidieť, korelácie sú vysoké a blížia sa k 1. Ak by boli rovné 1 je zřejmé, že toto spôsobuje problémy, filter má v podstate dve rovnako sa meniace zložky, a nemá spôsob ako sa medzi nimi rozhodnúť. Dá sa preto predpokladať, že pri koreláciách takto blízkych 1 bude práve toto spôsobovať filtru ťažkosti. (Spolu s multikolinearitou, o ktorej môžeme na základe tejto korelačnej matice usudzovať a o ktorej bližšie pojednáva kapitola 5.)

Hodnoty kritérií pre tento prípad uvedieme v tabuľke 3.15.



Obr. 3.9: Graf Jensenovho alpha pre portfólio s 3 reálnymi výnosmi.

súčet štvorcov	4.15813
relatívny súčet štvorcov	0.00280576
počet nesprávnych odhadov	1237
relatívny počet nesprávnych odhadov	0.818122

Tabuľka 3.15: Hodnoty kritérií pre reálne portfólio s 3 zložkami.

3.5 Vplyvy na konvergenciu

Ako sme uviedli v predošlej podkapitole, je zrejmé, že v určitých prípadoch rozptyl Jensenovho alpha vplýva na rýchlosť konvergence Kalmanovho filtra. V tejto podkapitole sa pokúsime vysvetliť, prečo to tak je, a tiež ktoré ďalšie faktory vplývajú na presnosť výsledkov.

Ako prvé si vezmeme k preskúmaniu Jensenovo alpha. Pri umelých dátach v prvom ideálnom prípade výsledky výrazne neovplyvňuje, v príklade s reálnymi dátami už ale Jensenovo alpha vstrebáva zmeny v portfóliu. Aplikujeme preto filter na rovnaké portfólio s reálnymi dátami, v ktorom zafixujeme rozptyl inovácie Jensenovho alpha na nule.

Výsledné odhadnuté rozptyly uvádzame v tabuľke 3.17. Pre úplnosť uvádzame aj rozptyl inovácie Jensenovho alpha, hoci ten filter neodhadoval, ale bol mu zadán ako 0.

počet zložiek portfólia	3
rozdelenie náh. prechádzky	$N(0, 0.005^2)$
rozdelenie pozorovacieho šumu	$N(0, 0.005^2)$
Jensenovo alpha zahrnuté	nie
výnosy portfólia	reálne

Tabuľka 3.16: Parametre príkladu 6 pre inovácie reálneho portfólia bez Jensenovho alpha.

σ_α^2	0
$\sigma_{x_1}^2$	$1.13 \cdot 10^{-5}$
$\sigma_{x_2}^2$	$2.00 \cdot 10^{-5}$
$\sigma_{x_3}^2$	$1.72 \cdot 10^{-5}$
σ_w^2	$2.54 \cdot 10^{-5}$

Tabuľka 3.17: Odhady rozptylov pre inovácie reálneho trojzložkového portfólia bez Jensenovho alpha.

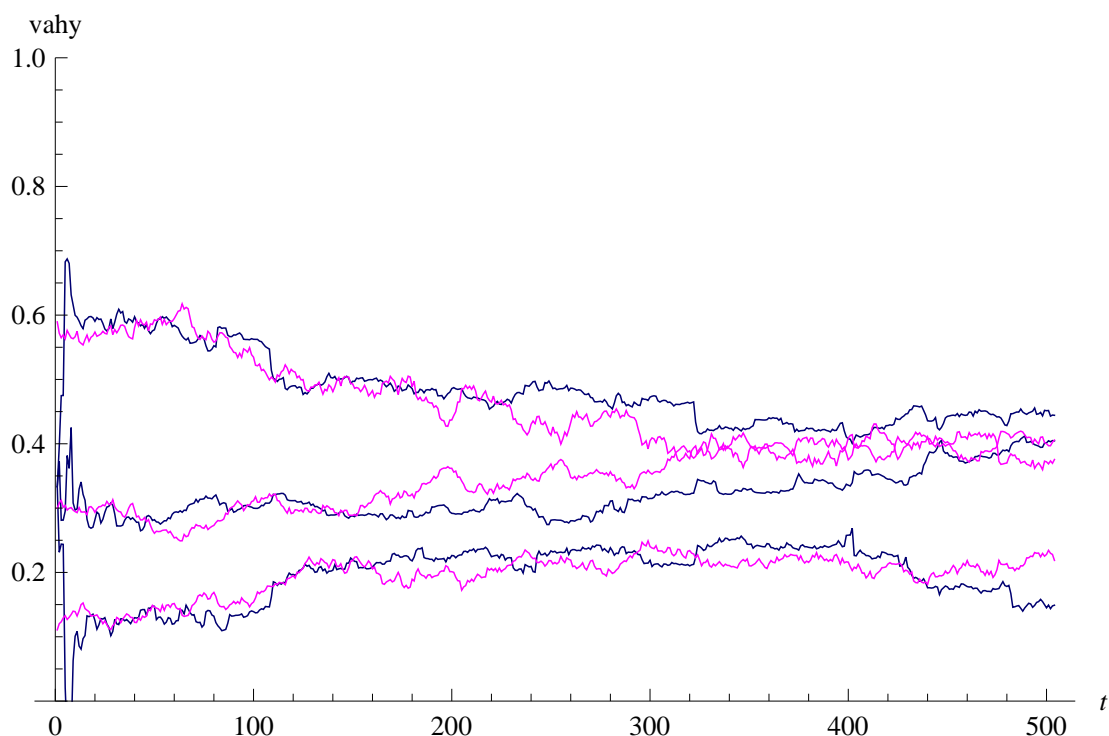
Odhady rozptylov sa oproti prípadu s Jensenovým alpha priblížili zadaným hodnotám. Rovnako došlo k spresneniu celkových výsledkov filtrácie, čo sa odráža na grafickom zobrazení výsledkov na obrázku 3.10, ako aj na hodnotách porovnávacích kritérií v tabuľke 3.18.

súčet štvorcov	2.13254
relatívny súčet štvorcov	0.00143896
počet nesprávnych odhadov	1224
relatívny počet nesprávnych odhadov	0.809524

Tabuľka 3.18: Hodnoty kritérií pre prípad reálneho trojzložkového portfólia bez Jensenovho alpha.

Z porovnania hodnôt kritérií s hodnotami pre portfólio so zahrnutým Jensenovým alpha vyplýva, že hoci sa počet nesprávnych odhadov vynechaním alpha znížil len o 13 (cca 0.9%), súčet štvorcov klesol skoro na polovicu. Z tohto vidieť, že vynechaním Jensenovho alpha z modelu je možné zlepšiť výsledky filtrácie.

Ďalším faktorom, vplývajúcim na presnosť konverencie, môže byť počet pozorovaní. Vráťme sa k ideálnemu prípadu s umelými dátami (príklad 1). Tentokrát ale vygenerujeme časový rad dĺžky 50 000. Použitá náhodná prechádza z príkladu 1 má síce veľký rozptyl pre praktické účely (pri 50 000 pozorovaniach sa váhy menia natoľko, že sa dostanú z intervalu (0,1)), pre presné porovnanie výsledkov ho ale ponecháme aj pre tento príklad. Použité parametre príkladu uvádzame v tabuľke 3.19



Obr. 3.10: Graf skutočných (ružové) a vyfiltrovaných váh pre portfólio s 3 reálnymi zložkami bez Jensenovho alpha.

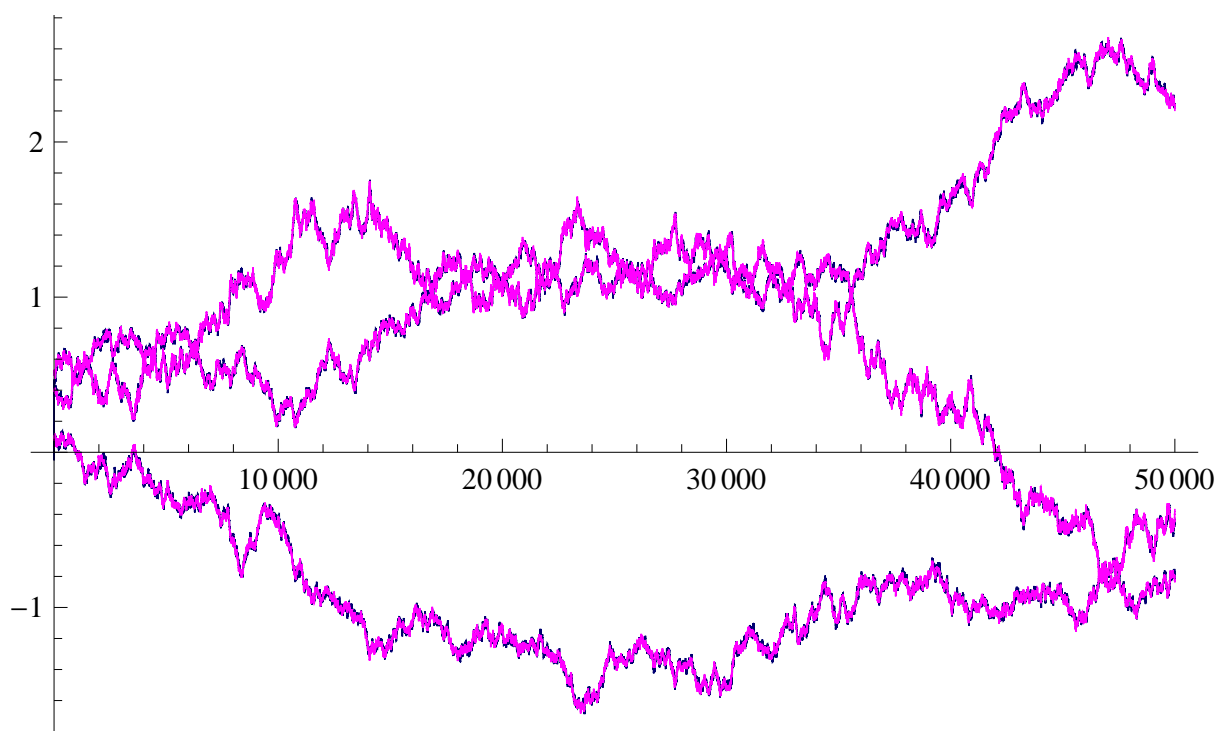
počet zložiek portfólia	3
dĺžka časového radu	50 000
rozdelenie náh. prechádzky	$N(0, 0.005^2)$
rozdelenie pozorovacieho šumu	$N(0, 0.005^2)$
Jensenovo alpha zahrnuté	áno
výnosy portfólia	generované z $N(0, 0.05^2)$

Tabuľka 3.19: Parametre príkladu 7 pre ideálny prípad a dlhšie časové rady.

Výsledky filtrácie uvádzame na grafe 3.11 a v tabuľke 3.20. Pripomeňme, že relatívne hodnoty kritérií sú spočítané z absolútnych, delených počtom pozorovaní (s vynechaním prvých 10 pozorovaní), v tomto prípade 149 970. Porovnaním hodnôt kritérií s hodnotami pre ideálny prípad, tabuľka 3.3, zisťujeme, že absolútny súčet štvorcov sa zmenšil takmer 100krát, zatiaľ čo absolútny počet nesprávnych odhadov sa zmenšil len o 0.005. Filtrácia sa pridaním pozorovaní zlepšuje v zmysle presnejších odhadov, t.j. odhady sú bližšie skutočným váham, hoci stále nie sú úplne presné.

Ďalším významným faktorom vplývajúcim na konvergenciu je počet zložiek portfólia. Začínali sme s 3 zložkami portfólia. Zvýšime tento počet na 8 a pozrieme sa, ako sa mení presnosť filtrácie pre reálne, aj umelo generované dáta.

Najskôr 8 zložiek umelo generovaných dát zložených do portfólia rovnako ako v ideálnom prípade. Parametre tohto príkladu uvádzame v tabuľke 3.21.



Obr. 3.11: Graf skutočných (ružové) a vyfiltrovaných váh (modré) pre portfólio s 3 reálnymi zložkami o dĺžke 50 000.

súčet štvorcov	92.6606
relatívny súčet štvorcov	0.000617737
počet nesprávnych odhadov	102120
relatívny počet nesprávnych odhadov	0.6808

Tabuľka 3.20: Hodnoty kritérií pre ideálny prípad s časovými radmi dĺžky 50 000.

Výsledky filtrácie pre tento príklad uvádzame v tabuľke 3.22.

Odhady rozptylov sú všetky menšie ako skutočné rozptyly, pozrieme sa preto na grafické výsledky na obrázku 3.12.

Z grafických výsledkov vidieť, že odhady nasledujú váhy, keďže ale grafické výsledky nie sú úplne prehľadné, najviac nám povedia hodnoty vypočítaných kritérií.

Napriek tomu, že sa celkový počet zložiek zvýšil hneď na 8, celkový súčet štvorcov nenarástol tak výrazne. Filtráciu preto môžeme považovať za presnú aj pre väčší počet zložiek. Poznamenajme ešte jeden výrazný aspekt v uvádzaných príkladoch, ktorým je čas výpočtu. Zatiaľ čo pri trojzložkovom portfóliu s časovými radmi dĺžky 504 trvá výpočet len pár sekúnd, pre časové rady dĺžky 50000 sa doba výpočtu výrazne predlžuje. Rovnako sa predlžuje doba výpočtu pri pridávaní zložiek.

Uveďme si teraz príklad s portfóliom s 8 reálnymi zložkami. Zložíme ho z ôsmich výnosov akcií stiahnutých zo stránok pražskej burzy. Ide o akcie spoločností - VIG, Erste Group Bank, Komerční Banka, CETV, UNIPETROL, Philip Morris ČR, Pegas Nonwovens a AAA. Parametre tohto príkladu uvádzame v tabuľke 3.24.

počet zložiek portfólia	8
rozdelenie náh. prechádzky	$N(0, 0.005^2)$
rozdelenie pozorovacieho šumu	$N(0, 0.005^2)$
Jensenovo alpha zahrnuté	áno
výnosy portfólia	generované z $N(0, 0.05^2)$

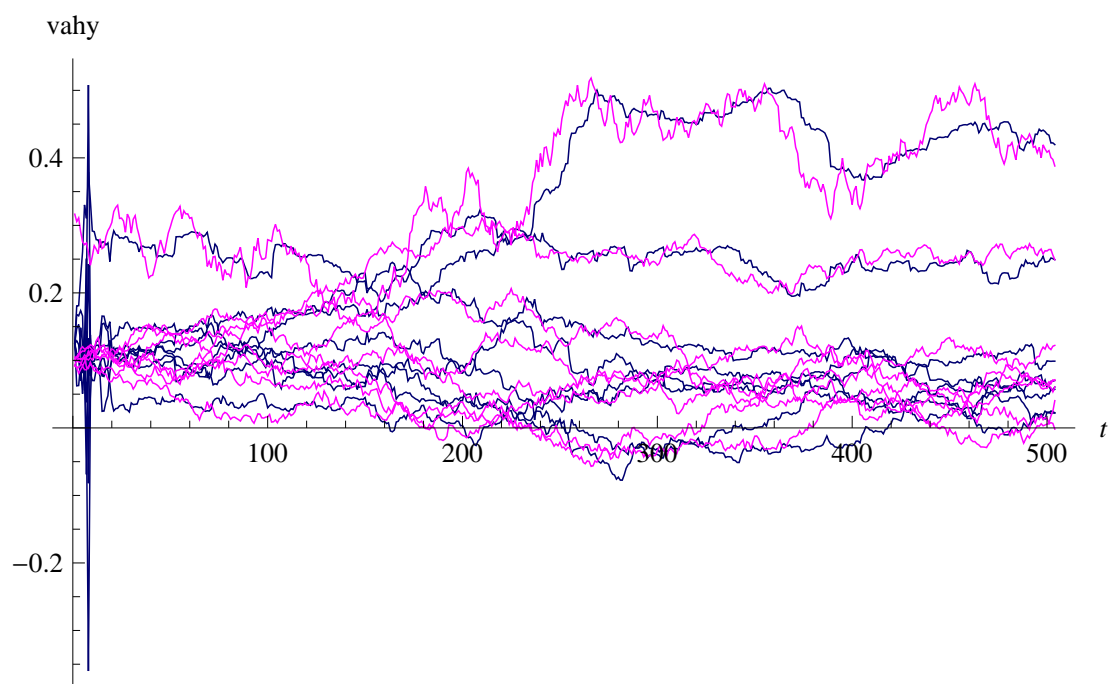
Tabuľka 3.21: Parametre príkladu 8 pre portfólio s 8 umelo generovanými zložkami.

σ_α^2	$1.51241 \cdot 10^{-5}$
$\sigma_{x_1}^2$	$1.66939 \cdot 10^{-5}$
$\sigma_{x_2}^2$	$1.65083 \cdot 10^{-5}$
$\sigma_{x_3}^2$	$1.63317 \cdot 10^{-5}$
$\sigma_{x_4}^2$	$1.57456 \cdot 10^{-5}$
$\sigma_{x_5}^2$	$1.59223 \cdot 10^{-5}$
$\sigma_{x_6}^2$	$1.56992 \cdot 10^{-5}$
$\sigma_{x_7}^2$	$1.59173 \cdot 10^{-5}$
$\sigma_{x_8}^2$	$1.93202 \cdot 10^{-5}$
σ_w^2	0

Tabuľka 3.22: Odhady rozptylov pre inovácie v portfóliu s 8 umelo generovanými zložkami.

Grafické výsledky pre toto portfólio opäť uvádzame, hoci viac nám povedia odhady rozptylov a napočítané hodnoty porovnávacích kritérií.

Hodnoty porovnávacích kritérií pre reálne portfólio s 8 zložkami uvádzame v tabuľke 3.26. V porovnaní s rovnakým portfóliom s umelo generovanými výnosmi sú hodnoty všetkých kritérií väčšie, pri porovnaní s reálnym portfóliom s 3 zložkami však nárast kritérií nie je výrazný.



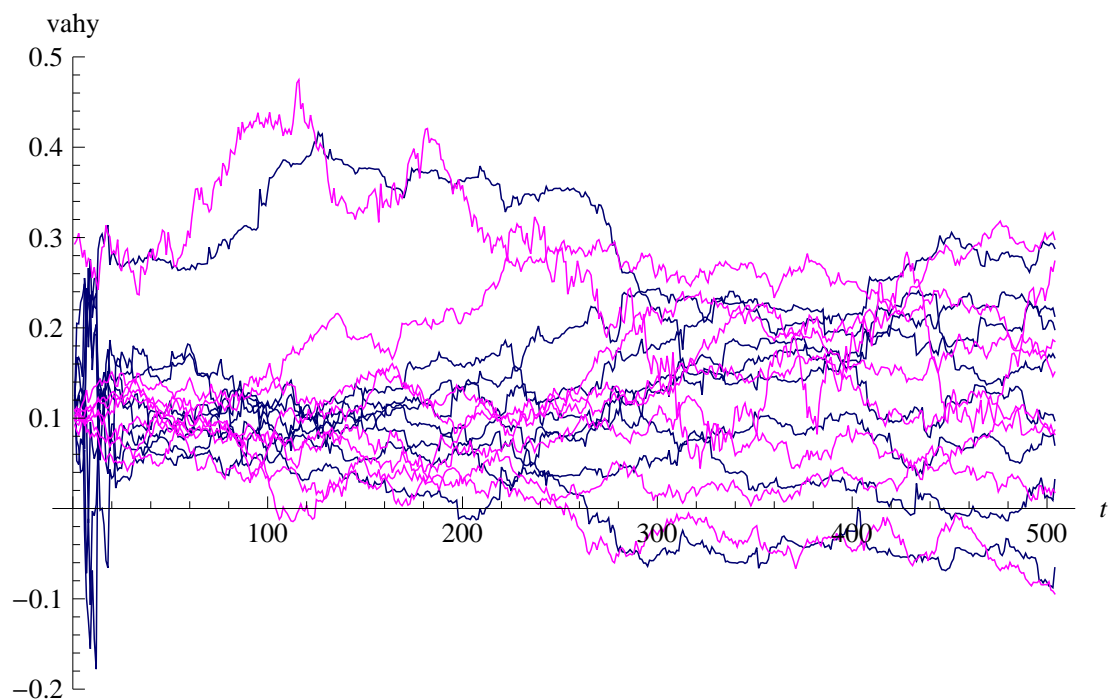
Obr. 3.12: Graf skutočných (ružové) a vyfiltrovaných váh pre portfólio s 8 umelo generovanými zložkami.

súčet štvorcov	2.45209
relatívny súčet štvorcov	0.000620469
počet nesprávnych odhadov	2709
relatívny počet nesprávnych odhadov	0.671875

Tabuľka 3.23: Hodnoty kritérií pre portfólio s 8 umelo generovanými zložkami.

počet zložiek portfólia	8
rozdelenie náh. prechádzky	$N(0, 0.005^2)$
rozdelenie pozorovacieho šumu	$N(0, 0.005^2)$
Jensenovo alpha zahrnuté	áno
výnosy portfólia	reálne

Tabuľka 3.24: Parametre príkladu 9 pre portfólio s 8 zložkami reálnych výnosov.



Obr. 3.13: Graf skutočných (ružové) a vyfiltrovaných váh (modré) pre portfólio s 8 reálnymi zložkami.

σ_{α}^2	$1.64583 \cdot 10^{-5}$
$\sigma_{x_1}^2$	$1.71611 \cdot 10^{-5}$
$\sigma_{x_2}^2$	$1.72121 \cdot 10^{-5}$
$\sigma_{x_3}^2$	$1.72611 \cdot 10^{-5}$
$\sigma_{x_4}^2$	$1.75768 \cdot 10^{-5}$
$\sigma_{x_5}^2$	$1.74801 \cdot 10^{-5}$
$\sigma_{x_6}^2$	$1.71321 \cdot 10^{-5}$
$\sigma_{x_7}^2$	$1.73223 \cdot 10^{-5}$
$\sigma_{x_8}^2$	$1.83669 \cdot 10^{-5}$
σ_w^2	0

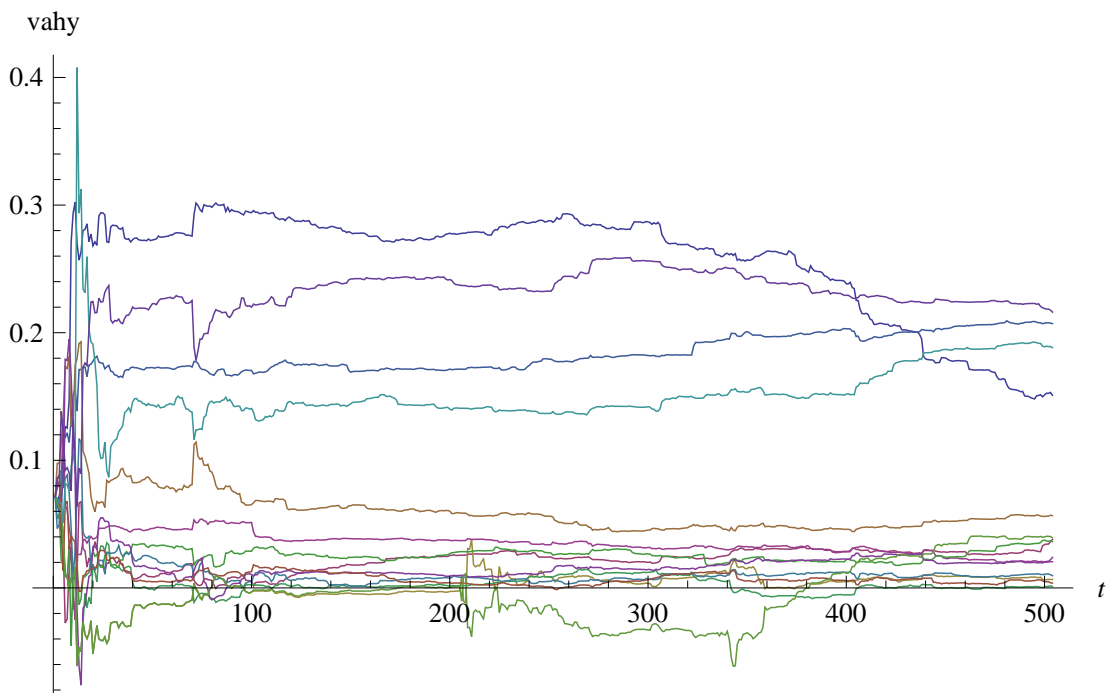
Tabuľka 3.25: Odhady rozptylov pre portfólio s 8 reálnymi zložkami.

súčet štvorcov	7.01601
relatívny súčet štvorcov	0.00177531
počet nesprávnych odhadov	3222
relatívny počet nesprávnych odhadov	0.799107

Tabuľka 3.26: Hodnoty kritérií pre portfólio s 8 reálnymi zložkami.

3.6 Reálny príklad PX indexu

Na koniec tejto kapitoly si ukážeme ešte jeden reálny príklad s reálnym portfóliom. V tomto prípade sme stiahli zo stránky pražskej burzy ceny 14 akcií, z ktorých sa počíta index pražskej burzy PX. Napočítali sme si výnosy týchto akcií a výnosy PX indexu a vyfiltrovali váhy jednotlivých zložiek pre výpočet PX. Tento reálny príklad plne zodpovedá doteraz modelovaným simuláciám. Váhy zložiek tvoriacich PX index sa menia na dennej báze, hoci minimálne. Pozrime sa teraz na grafické výsledky filtrácie.



Obr. 3.14: Grafy vyfiltrovaných váh zložiek PX indexu.

Ako vidieť z obrázku 3.14, Kalmanov filter odhadoval jednu zložku PX indexu pre isté časy zápornú, čo spôsobilo, že následne iná zložka má váhu vyššiu ako by mala mať - podľa pravidiel pražskej burzy totiž nesmie mať žiadna zložka váhu viac ako 20%, od marca 2011 viac ako 25% (údaje sú za rok 2010 a 2011).

Ku koncu uvažovaného obdobia sa však už váhy približujú číslam pražskej burzy. Zloženie PX indexu k 30.12.2011 je uvedené v tabuľke 3.27.

Ako vidieť z tabuľky 3.27, PX index sa skladá z niekoľkých významných zložiek (napr. ČEZ s váhou viac ako 20%) a niekoľkých menej významných zložiek (pod 1%). Nás spravidla zaujímajú tie najvýznamnejšie zložky, zatiaľ čo informáciu o tom, že v portfóliu je 0.14% nejakej akcie nepovažujeme za tak dôležitú. Vyskúšame preto, či nepomôže vylepšiť výsledky filtra, ak by sme vynechali tie menej dôležité zložky a pokúsili sa mu tým uľahčiť výpočet.

Na obrázku 3.15 vidíme vyfiltrované váhy portfólia, pre zjednodušenie úlohy odobraním zložiek. Zníženie počtu zložiek Kalmanovmu filtru nijako nepomohlo, vidieť, že výsledky sú podobné ako predtým. Filter predpokladá vyššie váhy ako reálne sú pre niektoré akcie a na druhej strane príliš nízke váhy pre iné akcie.

Název	ISIN	Váha [%]
ČEZ	CZ0005112300	20,3
ERSTE GROUP BANK	AT0000652011	17,76
KOMERČNÍ BANKA	CZ0008019106	16,92
TELEFÓNICA C.R.	CZ0009093209	16,5
VIG	AT0000908504	13,69
NWR	GB00B42CTW68	4,8
UNIPETROL	CZ0009091500	4,15
PHILIP MORRIS ČR	CS0008418869	3,22
CETV	BMG200452024	0,97
FORTUNA	NL0009604859	0,64
PEGAS NONWOVENS	LU0275164910	0,56
ORCO	LU0122624777	0,19
AAA	NL0006033375	0,16
KITD	US4824702009	0,14
Celkem		100

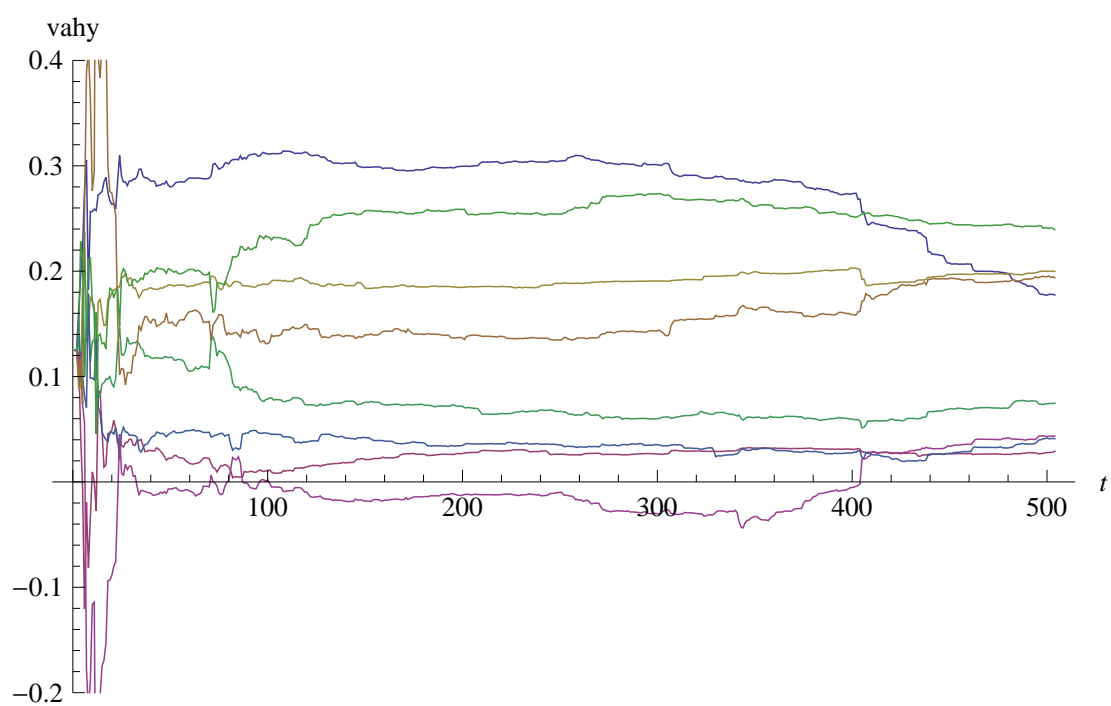
Tabuľka 3.27: Zloženie PX indexu k 30.12.2011.

Pozrime sa preto ešte na korelačnú maticu týchto ôsmich akcií.

$$\begin{pmatrix} 1. & 0.8978 & 0.8503 & 0.2085 & -0.6574 & -0.3261 & 0.3427 & 0.5894 \\ 0.8978 & 1. & 0.8609 & 0.1487 & -0.5618 & -0.4486 & 0.308 & 0.5922 \\ 0.8503 & 0.8609 & 1. & 0.2595 & -0.3562 & -0.2906 & 0.1204 & 0.3891 \\ 0.2085 & 0.1487 & 0.2595 & 1. & -0.3866 & -0.2022 & -0.317 & -0.1013 \\ -0.6574 & -0.5618 & -0.3562 & -0.3866 & 1. & 0.3683 & -0.3747 & -0.6125 \\ -0.3261 & -0.4486 & -0.2906 & -0.2022 & 0.3683 & 1. & -0.0459 & -0.1245 \\ 0.3427 & 0.308 & 0.1204 & -0.317 & -0.3747 & -0.0459 & 1. & 0.6888 \\ 0.5894 & 0.5922 & 0.3891 & -0.1013 & -0.6125 & -0.1245 & 0.6888 & 1. \end{pmatrix}$$

Napriek tomu, že predpoklady Kalmanovho filtra nepožadujú nekorelovanosť, je zrejmé, že v prípade reálnych dát, ktoré sa od umelých líšia práve nenulovou koreláciou, nespĺňajú výsledky naše očakávania.

V ďalších kapitolách si ukážeme, ako zvláda rolovacia regresia dynamickú analýzu portfólia a pokúsime sa vylepšiť obe metódy tak, aby dávali uspokojivé výsledky.



Obr. 3.15: Grafy vyfiltrovaných váh PX indexu pre 8 akcií.

4. Rolovacia regresia

V tejto kapitole sa zameriame na popis rolovacej regresie, (v anglickej literatúre nazývanej rolling príp. locally weighted regression), ako inej metódy k dynamickej analýze zloženia portfólia. Teória v tejto kapitole sa opiera o [7].

4.1 Teoretický základ

Rolovacia analýza modelov časových radov sa často používa k odhadnutiu stability modelu v čase. Pri analyzovaní finančných časových radov sa ako základný predpoklad používa konštantnosť modelu v čase. V skutočnosti sa ale ekonomické prostredie často mení a preto nie je rozumné predpokladať, že parametre modelu sú konštantné. Bežnou technikou pre posúdenie konštantnosti parametrov je odhad parametrov cez rolovacie okno pevnej šírky. Ak sú parametre skutočne konštantné cez celú vzorku, potom by sa odhady parametrov pre jednotlivé okná nemali veľmi líšiť. Ak sa parametre v nejakej časti vzorku menia, potom by rolovacie odhady mali zaznamenať túto nestabilitu.

Rolovacia analýza sa často používa napríklad k backtestovaniu štatistických modelov na historických dátach kvôli ohodnoteniu stability a presnosti predikcie.

Jednou z používaných metód rolovacej analýzy je metóda rolovacej regresie. Rolovacia regresia typicky používa rovnomerné rozloženie váh po celej šírke rolovacieho okna, alebo vhodným spôsobom klesajúce rozloženie váh pozorovaní od stredu rolovacieho okna smerom ku krajom. My sa zameriame na klesajúce váhy pozorovaní.

4.2 Lokálne vážená regresia

V anglickej literatúre sa objavuje pod názvom Locally weighted regression (v českej ako lokálne vážená regresia, prípadne rolovacia regresia). Ide o upravenú lineárnu regresiu, ktorá kladie väčší význam pozorovaniam blízko odhadovaného pozorovania, proti vzdialenejším hodnotám.

Zopakujme si najskôr, ako vyzerá model lineárnej regresie, ktorý potom následne rozšírime na lokálne váženú regresiu.

Buďte Y_1, \dots, Y_n nezávislé náhodné veličiny so strednými hodnotami

$$\mathbb{E}Y_i = \beta_1 + \beta_2 x_i \quad i = 1, \dots, n$$

a so spoločným rozptylom $\text{var}Y_i = \sigma^2$, kde $\beta_j \in \mathbb{R}$ a $\sigma^2 \geq 0$ pre všetky j sú neznáme parametre zatiaľ čo x_i sú známe konštanty. Popísaný model zodpovedá situácii, kedy medzi veličinami x a y existuje lineárna závislosť $y = \beta_1 + \beta_2 x$, ktorej parametre ale nepoznáme a informácie o nich získavame len experimentálne, totiž tak, že pre zvolné hodnoty x_i nameriame príslušné hodnoty y_i zaťažené chybou. Výsledky merania sú teda $Y_i = \beta_1 + \beta_2 x_i + e_i$, kde e_i sú chyby. Ak sú tieto chyby nezávislé, náhodné, bez systematickej zložky, čo vyjadríme ako $\mathbb{E}e_i = 0$ a ak sú merania rovnako presné, čo vyjadríme požiadavkou $\text{var}e_i = \sigma^2$ dostávame náš model. Cieľom tohoto cvičenia je odhadnúť parametre modelu β_1, β_2 metódou najmenších štvorcov.

Odhady parametrov sú potom také hodnoty, ktoré minimalizujú súčet

$$\sum_{i=1}^n (Y_i - \beta_1 - \beta_2 x_i)^2.$$

Pre váženú regresiu je myšlienka rovnaká, zmena je v záverečnej sume, ktorú minimalizujeme. Zatiaľ čo v lineárnej regresii sú členy sumy kvadratické odchylky, u vázenej regresie im ešte pridáme váhy. Úloha sa potom mení na úlohu nájsť parametre β_i tak, aby sme minimalizovali

$$\sum_{i=1}^n w_i (Y_i - \beta_1 - \beta_2 x_i)^2.$$

Váhy w_i je možné voliť niekoľkými spôsobmi. Veľmi rozšírený spôsob je voliť váhy klesajúce od stredu ku krajom rolovacieho okna. Vykreslené na grafe tieto váhy pripomínajú hustotu normálneho rozdelenia. w_i majú potom tvar

$$w_i = \exp\left\{-\frac{(x_i - x)^2}{2\tau^2}\right\},$$

kde x je stred rolovacieho okna a τ je parameter ovplyvňujúci rýchlosť klesania váh od stredu rolovacieho okna.

Pre takto zvolené váhy platí, že pre malé $|x_i - x|$ sú váhy blízke 1 (t.j. pre hodnoty blízko odhadovaného pozorovania sú váhy významné) a pre veľké $|x_i - x|$ sú váhy blízke 0 (t.j. pre hodnoty vzdalujúce sa od odhadovaného pozorovania sú váhy zanedbateľné).

Prevedieme túto teóriu na naše značenie a náš problém. Majme portfólio s hodnotami y_t v čase t . Pre jednoduchosť predpokladajme na začiatok, že máme len dve zložky portfólia a teda potrebujeme odhadnúť len dve váhy. V čase t má naša rovnica pozorovaní tvar

$$y_t = x_{1t}\beta_{1t} + x_{2t}\beta_{2t} + \epsilon_t,$$

kde y_t je hodnota známa v čase t a x_{1t} a x_{2t} sú tiež známe. Parametre, ktoré potrebujeme odhadnúť, sú β_{1t} a β_{2t} . Metódou najmenších štvorcov teda hľadáme minimum zo sumy

$$\sum_{i=m}^n w_i (y_i - \beta_{1i}x_{1i} - \beta_{2i}x_{2i})^2,$$

kde m a n označujú začiatok a koniec rolovacieho okna.

4.3 Praktické aplikácie

Podobne ako pri skúmaní Kalmanovho filtra, sa teraz pozrieme na praktické využitie rolovacej regresie a na presnosť požadovaných výsledkov.

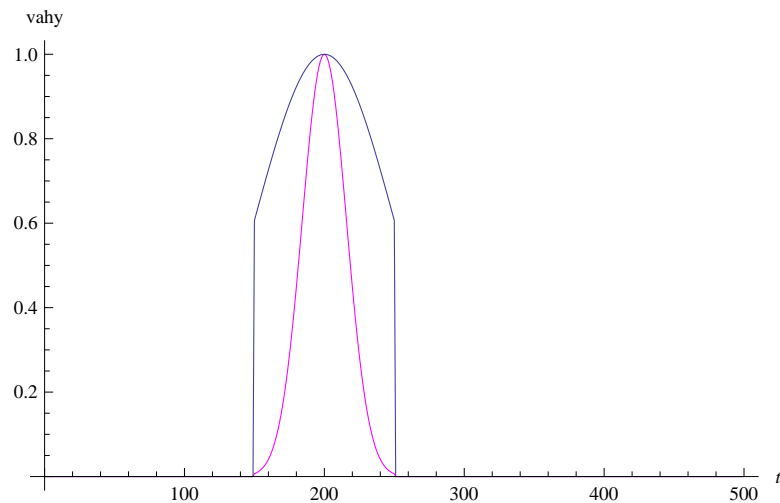
4.3.1 Portfólio s umelo generovanými výnosmi

Opäť začneme jednoduchým príkladom umelého portfólia, aby sme preskúmali, ako vlastne nastaviť parametre rolovania - váhy a šírku rolovacieho okna. Vygenerujeme si výnosy z normálneho rozdelenia, tak ako v ideálnom prípade pre Kalmanov filter, rovnako vygenerujeme aj váhy portfólia (slovo váhy odteraz používame nielen pre označenie váh portfólia ale aj pre váhy pozorovaní v rolovacej regresii, z kontextu by však vždy malo byť zrejmé, ktoré váhy máme na mysli) pridáme k nim náhodnú prechádzku, a nakoniec pozorovací šum. Zvolené parametre uvádzame v tabuľke 4.1.

počet zložiek portfólia	3
rozdelenie náh. prechádzky	$N(0, 0.005^2)$
rozdelenie pozorovacieho šumu	$N(0, 0.005^2)$
Jensenovo alpha zahrnuté	nie
výnosy portfólia	generované z $N(0, 0.05^2)$

Tabuľka 4.1: Parametre ideálneho prípadu pre regresiu.

Pozrieme sa teraz, ako nastavenie parametra τ ovplyvňuje váhy zvolené k lokálnej váženej regresii. Metóda, ktorú sme vyššie určili pre výpočet váh, dáva väčšiu váhu pozorovaniam blízko stredu okna a menšiu váhu tým vzdialenejším. Tvarom rozdelenie váh pripomína hustotu normálneho rozdelenia a τ ovplyvňuje rýchlosť klesania váh od stredu rolovacieho okna. Na obrázku 4.1 vidieť dve funkcie váh. Ružová čiara sú váhy pre $\tau = 500$, modrá čiara pre $\tau = 5000$. V uvedenom príklade je šírka rolovacieho okna volená ako 100 pozorovaní a zobrazené váhy sú pre stred rolovacieho okna 200. Vidíme teda, že pri $\tau = 5000$ klesajú váhy pozorovaní v okne postupne až k 0.6 pre krajné pozorovania (150 a 250) a ďalej už sú nulové. Pre $\tau = 500$ klesajú váhy rýchlejšie a v zvolenom okne dokonvergujú až k nule, mimo okna pozorovaní sú nulové.

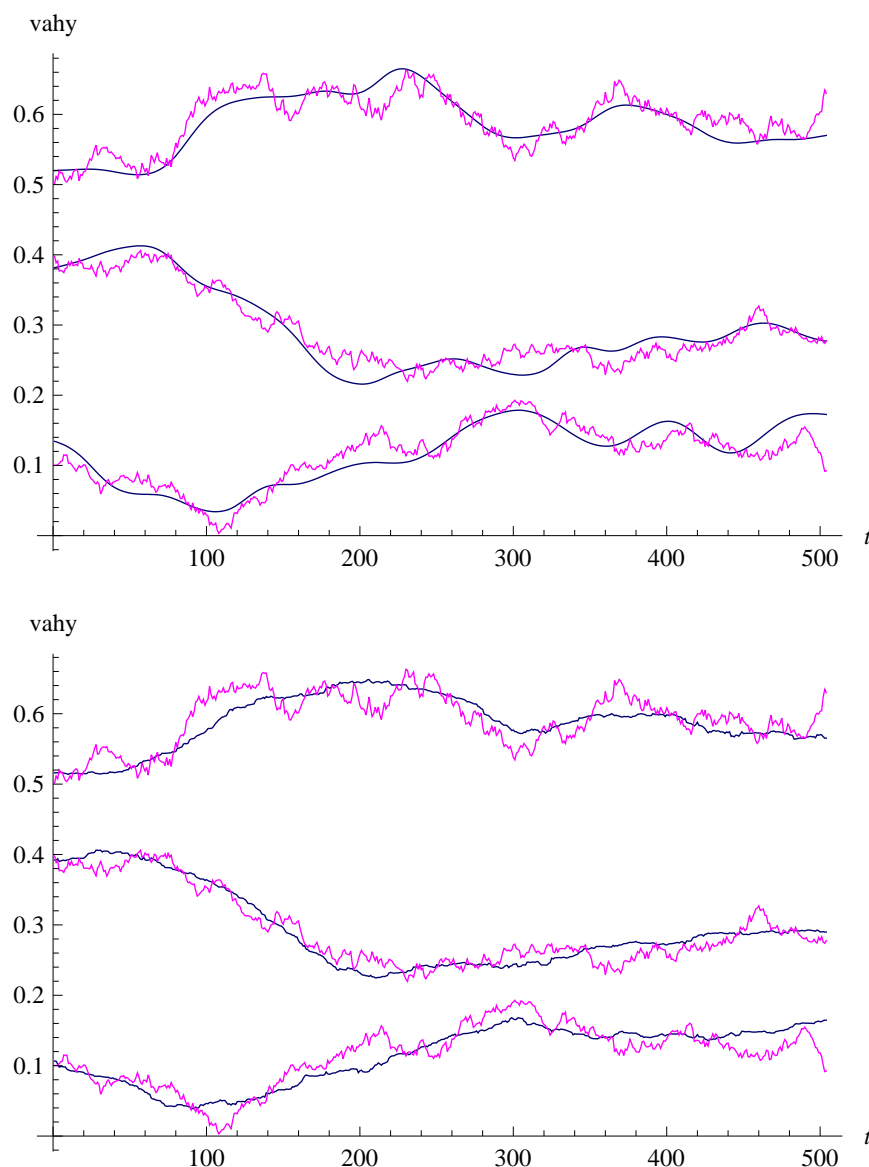


Obr. 4.1: Váhy regresie v závislosti na τ .

Zaujímá nás, ako sa pre takto zvolené váhy líšia výsledky regresie pre vyššie popísaný príklad umelého portfólia. Na obrázku 4.2 máme hore odhadnuté váhy portfólia pre $\tau = 5000$, dole pre $\tau = 500$ (šírka okna je 100).

Z grafického zobrazenia výsledkov nie je zrejmé, ktorá voľba parametrov je lepšia. Vidieť však rozdiel vo vyfiltrovaných odhadoch. Pre $\tau = 500$ sú vyfiltrované váhy portfólia hladké čiary, pre $\tau = 5000$ nie sú. Toto je spôsobené nespojitosťou váh rolovacej regresie pri posúvaní rolovacieho okna pre $\tau = 5000$.

Pozrieme sa ešte na hodnoty porovnávacích kritérií, uvedených v tabuľke 4.2. Aj v tejto kapitole budeme používať súčet štvorcových odchýlok a počet nesprávnych odhadov. Definícia týchto kritérií zostáva rovnaká, ako sme zaviedli v kapitole 3 pre Kalmanov filter, aby bolo možné výsledky porovnávať.



Obr. 4.2: Odhadnuté váhy umelého portfólia hore pre $\tau = 500$, dole pre $\tau = 5000$ pomocou rolovacej regresie.

Z výsledkov sa zdá, že by správne zvolená rolovacia regresia mohla ponúknuť rovnako kvalitné výsledky ako Kalmanov filter. Pokúsime sa preto vyriešiť aj zložitejšie príklady a porovnať výsledky regresie s Kalmanovým filtrom.

4.3.2 Portfólio s reálnymi výnosmi

Na začiatok sa pozrime ako zvláda rolovacia regresia reálne dáta. Použijeme reálne výnosy (stiahnuté zo stránky pražskej burzy, konkrétne výnosy Erste Group, VIG a Komerční banky), z ktorých zložíme umelé portfólio, využitím parametrov ideálneho príkladu uvádzaných v tabuľke 4.3.

Pre dĺžku okna 100 a $\tau = 500$ (hore) a $\tau = 5000$ (dole) vidíme výsledky

	$\tau = 500$	$\tau = 5000$
súčet štvorcov	0.580167	0.607547
relatívny súčet štvorcov	0.000383708	0.000401817
počet nesprávnych odhadov	963	954
relatívny počet nesprávnych odhadov	0.636905	0.630952

Tabuľka 4.2: Hodnoty kritérií pre ideálny prípad a $\tau = 500$ a 5000 .

na obrázku 4.3.

Tieto už nie sú tak presné ako pre ideálny prípad, čo sa odráža aj na hodnotách kritérií v tabuľke 4.4. Nepresnosti sú spôsobené významnou koreláciou dát, ktorá sťažuje regresii filtráciu. Preto je pri reálnych dátach dôležitejšie ako pri umelých správne nastavenie parametrov regresie.

počet zložiek portfólia	3
rozdelenie náh. prechádzky	$N(0, 0.005^2)$
rozdelenie pozorovacieho šumu	$N(0, 0.005^2)$
Jensenovo alpha zahrnuté	nie
výnosy portfólia	reálne

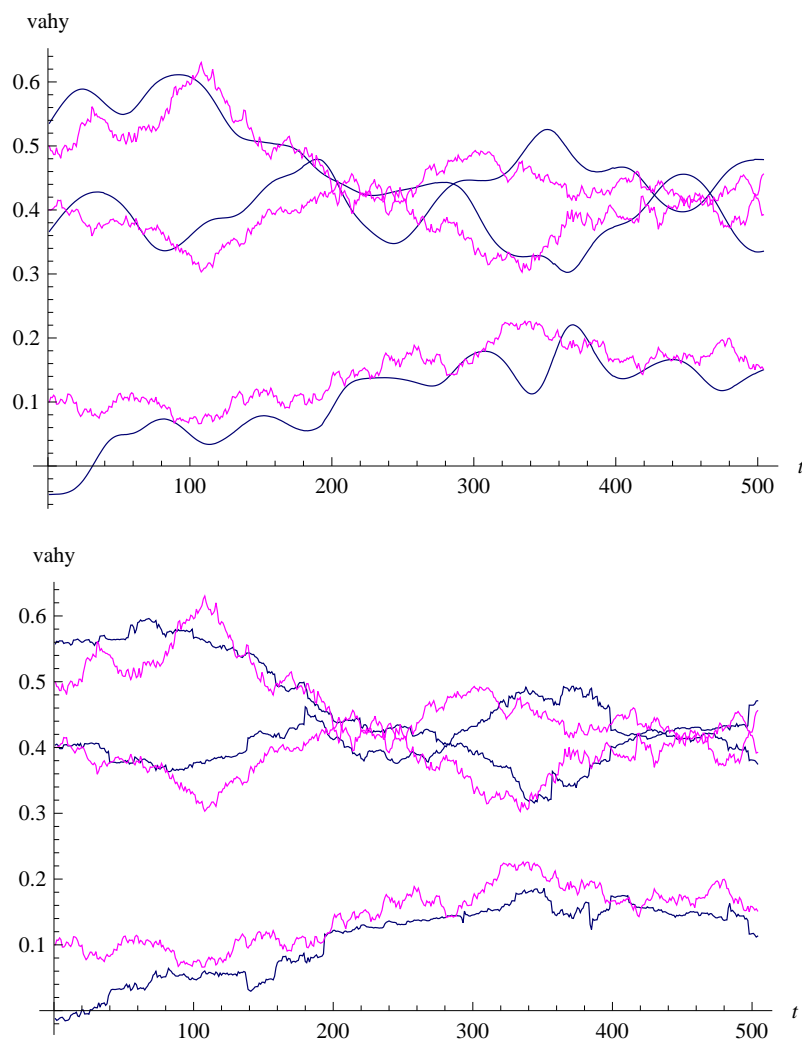
Tabuľka 4.3: Parametre ideálneho prípadu s reálnym portfóliom pre regresiu.

	$\tau = 500$	$\tau = 5000$
súčet štvorcov	3.17857	2.36619
relatívny súčet štvorcov	0.00210223	0.00156494
počet nesprávnych odhadov	1289	1280
relatívny počet nesprávnych odhadov	0.852513	0.846561

Tabuľka 4.4: Hodnoty kritérií pre reálne portfólio a regresiu.

Zameriame sa na nájdenie správneho nastavenia parametrov regresie pre čo najpresnejšie výsledky. Vplyv parametrov si ukážeme na jednoduchom príklade. Zostavme portfólio z reálnych výnosov (keďže pre tieto vyžaduje nastavenie parametrov viac presnosti) s váhami, ktoré budú po častiach konštantné s občasnými skokovými zmenami. Príklad takéhoto portfólia, resp. už jeho regresiou odhadnutých váh, uvádzame na grafe 4.4.

Zaoberajme sa len zložkou 1 na grafe 4.4. My vieme, že k zmene váhy došlo v čase 90. Regresii sme ale poskytli okno šírky 50, čo znamená, že už v čase 65 zachytí zmenu na konci rolovacieho okna (z 0.1 na 0.2) a začne k nej konvergovať.



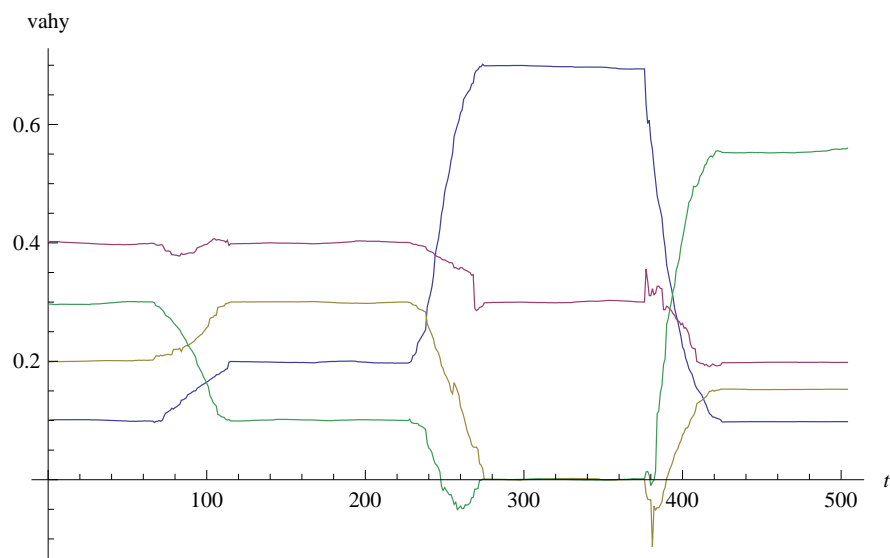
Obr. 4.3: Odhadnuté váhy reálneho portfólia s 3 zložkami, hore pre $\tau = 500$, dole pre $\tau = 5000$ pomocou rolovacej regresie.

Čím viac sa k zmene približuje, tým väčšie váhy majú tieto pozorovania a tým viac ovplyvňujú výsledný odhad. Ale až do času 115 je tento odhad ešte ovplyvňovaný „starou váhou“ t.j. 0.1. Poskytnutím širokého rolovacieho okna sme vlastne dosiahli to, že po celú dĺžku okna (50 dní) nemáme presný výsledok.

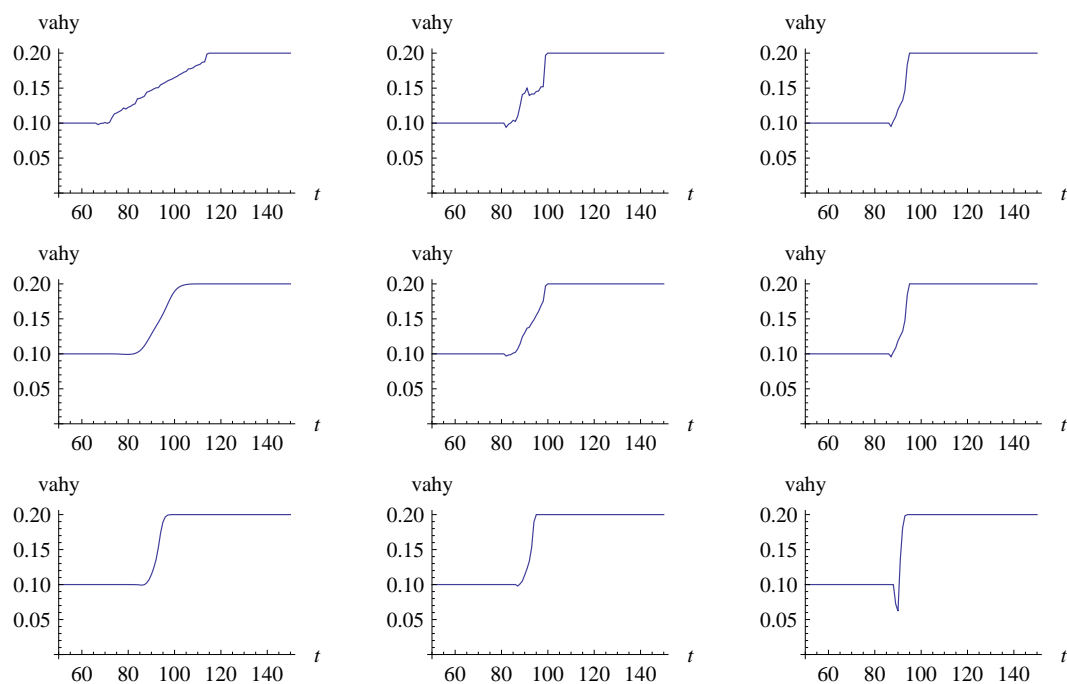
Zameriame sa preto na preskúmanie, aký vplyv má zvolená šírka okna a τ na odhad. Kvôli názornosti vezmeme len odhad prvej zložky a to od času 50 po čas 150.

Na obrázku 4.5 máme odhad pre rôzne kombinácie τ a šírky rolovacieho okna. Prvý riadok grafov sú odhady pre $\tau = 500$ a šírku okna postupne (zľava do prava) 50, 20 a 10. V druhom riadku sú odhady pre $\tau = 50$ a šírku okna postupne 50, 20 a 10. V poslednom riadku sú odhady pre $\tau = 10$ a šírku okna 5, $\tau = 10$ a šírku okna 10 a $\tau = 2$ a šírku okna 50.

Cieľom bolo ukázať, že s klesajúcim τ a klesajúcou šírkou okna sa znižuje interval, na ktorom metóda konverguje k správne výsledku. Zároveň však pre niektoré kombinácie už nie je podstatná šírka okna - τ dáva váhy len pre menšie okno pozorovaní ako je zadané rolovacie okno. Posledný obrázok zase ukazuje, že pre príliš malé τ sú výsledky nestabilné.



Obr. 4.4: Váhy reálneho portfólia s po častiach konštantnými váhami so skokmi pomocou rolovacej regresie pre $\tau = 500$ a šírku okna 50.



Obr. 4.5: Presnosť konverencie pre rôzne kombinácie parametrov modelu.

4.4 Optimalizácia lokálne váženej regresie

V predošlej podkapitole sme si ukázali, ako sa správa rolovacia regresia pre rôzne parametre τ a šírku okna. Ako ale voliť tieto parametre? Aby sme mohli zodpovedať túto otázku, potrebujeme zvoliť kritérium, pomocou ktorého určíme, či výsledky považujeme za kvalitné alebo nie. Pre tento účel využijeme súčet štvorcov definovaný už v kapitole Kalmanov filter, ale rovnako dobre poslúžia aj pri využití regresie.

Spočítame súčet kvadratických odchýliek pre niekoľko kombinácií parametrov.

τ	500	500	500	50	50	50
šírka rol. okna	50	20	10	50	20	10
$\sum_{i=1}^n (\hat{x}_i - x_i)^2$	6.66132	4.40466	4.2714	3.95111	3.80885	4.33272

τ	10	10	2
šírka rol. okna	50	10	50
$\sum_{i=1}^n (\hat{x}_i - x_i)^2$	3.81491	5.06074	16.3388

Tabuľka 4.5: Hľadanie optimálnej kombinácie parametrov pre prípad s po častiach konštantnými váhami so skokmi.

Tieto výsledky presne kopírujú výsledky z obrázku 4.5 (prvá kombinácia parametrov presne zodpovedá prvému obrázku, druhá druhému atď.). Potvrdil sa nám predpoklad, že kombinácia parametrov z posledného obrázku dáva už nestabilné výsledky, a tiež sa ukázalo, že rôzne kombinácie môžu dať rôzne presné výsledky.

Pozn.: Na obrázku 4.5 je zobrazená len časť výsledkov pre odhad jednej váhy v čase od 50 do 150 pre portfólio so štyrmi zložkami reálnych výnosov z obrázku 4.4. Súčet štvorcov je však počítaný pre kompletne výsledky, teda pre čas od 1 do 504 (dĺžka časového radu) a pre všetky zložky portfólia.

Z vyššie uvedených výsledkov vyplýva, že najlepšie výsledky dáva regresia s parametrami $\tau = 50$ a šírkou okna 20. Tieto parametre zabezpečujú najpresnejšie výsledky (z testovaných kombinácií parametrov). Je nutné poukázať, že takto zvolené parametre eliminujú vplyv už aj relatívne blízkych pozorovaní. To je spôsobené tým, že simuláciu sme spravili na príklade po častiach konštantných váh so skokmi. Výsledkom je, že sme dostali ako optimálne veľmi úzke okno, ktoré zachytí skok, ale nenechá sa ním dlho ovplyvňovať.

Avšak pre príklady aké sme používali doteraz, váhy zložiek portfólia meniace sa vplyvom náhodnej prechádzky, môžu byť vhodné iné parametre. Zopakujeme preto simuláciu pre iný prípad. Zostavme portfólio s tromi zložkami reálnych výnosov, kde váhy sa menia vplyvom náhodnej prechádzky generovanej z rozdelenia $N(0, 0.005^2)$. Výsledky tejto simulácie uvádzame v tabuľke 4.6.

τ	5000	5000	5000	500	500	500
šírka rol. okna	50	20	10	50	20	10
$\Sigma_{i=1}^n (\hat{x}_i - x_i)^2$	2.87804	4.08761	10.6266	2.98173	4.12739	10.6715

τ	50	50	50	10	10	2
šírka rol. okna	50	20	10	50	10	10
$\Sigma_{i=1}^n (\hat{x}_i - x_i)^2$	9.84301	9.84663	12.9644	34.189	34.1964	178.107

Tabuľka 4.6: Hľadanie optimálnej kombinácie parametrov pre prípad s váhami meniacimi sa náhodnou prechádzkou.

Pozn.: Tento spôsob nájdenia vhodného nastavenia parametrov je v praxi ťažko použiteľný. V prípade, že máme nejaké informácie o tom, ako dáta vyzerajú, je možné nagenerovať dáta rovnakých vlastností a na nich spraviť simuláciu nastavenia parametrov. Ak však nevieme o filtrovaných dátach, nič je prakticky nemožné odhadnúť, aké nastavenie parametrov je správne.

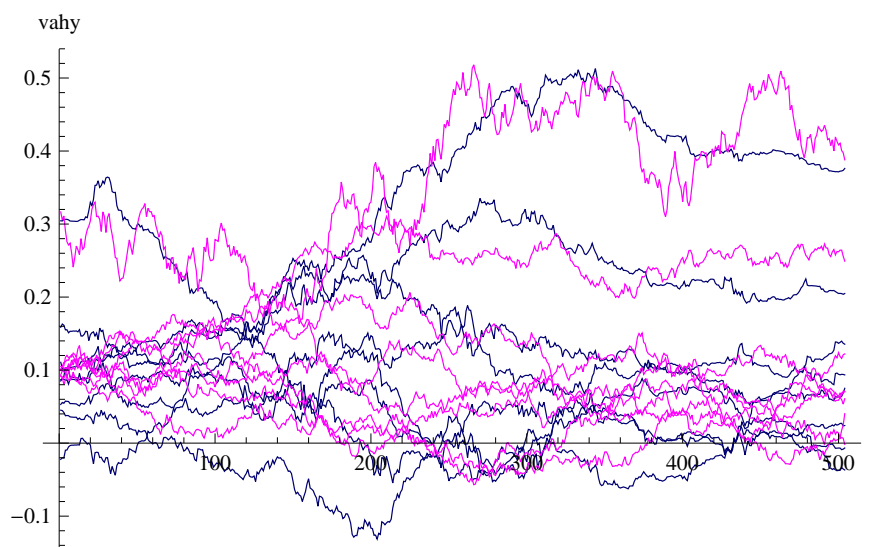
4.5 Praktické aplikácie - zložitejšie príklady

S využitím poznatkov o vhodnom nastavení parametrov lokálne váženej regresie sa teraz môžeme zaoberať aj zložitejšími príkladmi. Vieme už, že regresia funguje dobre na portfólio s tromi zložkami, zameriame sa preto na rozšírenie počtu zložiek. Pre možnosť neskoršieho porovnania výsledkov regresie s výsledkami Kalmanovho filtra sa pozrieme na rovnaký príklad, aký sme riešili už pomocou Kalmanovho filtra. Šlo o prípad portfólia zloženého z 8 zložiek, ktorých výnosy sme generovali z normálneho rozdelenia. Presné parametre tohto príkladu uvádzame v tabuľke 4.7.

počet zložiek portfólia	8
rozdelenie náh. prechádzky	$N(0, 0.005^2)$
rozdelenie pozorovacieho šumu	$N(0, 0.005^2)$
Jensenovo alpha zahrnuté	nie
výnosy portfólia	generované z $N(0, 0.05^2)$

Tabuľka 4.7: Parametre príkladu 8 zložkového umelého portfólia riešeného regresiou.

Pre úplnosť uvádzame na obrázku 4.6 aj grafické výsledky, hoci viac nám povedia napočítané hodnoty porovnávacích kritérií v tabuľke 4.8. Z týchto vidíme, že súčet štvorcov vzrástol oproti jednoduchším príkladom, stále ale nie je tak veľký, aby svedčil o nepoužiteľnosti metódy.



Obr. 4.6: Váhy umelého portfólia s 8 zložkami pomocou rolovacej regresie pre $\tau = 5000$ a šírku okna 100.

súčet štvorcov	10.3869
relatívny súčet štvorcov	0.00257613
počet nesprávnych odhadov	3341
relatívny počet nesprávnych odhadov	0.828621

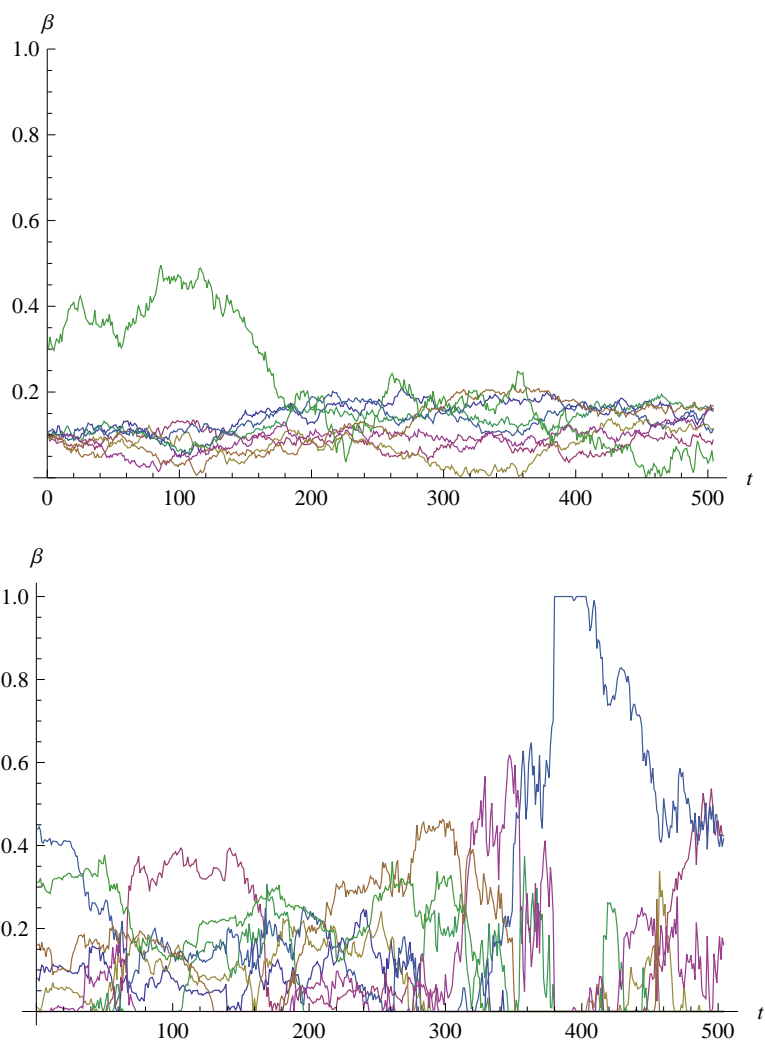
Tabuľka 4.8: Hodnoty kritérií pre umelé portfólio s 8 zložkami odhadované regresiou.

Podobne ako pre Kalmanov filter sme si potvrdili, že v prípade umelých dát dáva metóda dobré výsledky (pri správnom nastavení, ktoré môže byť v praxi ťažko odhadnuteľné, viz. poznámka na konci podkapitoly 4.4). Pozrieme sa ešte ako si poradí s reálnymi dátami.

Použijeme rovnaký príklad s ôsmimi zložkami ako v kapitole Kalmanov filter (podkapitola 3.5), kde sme zložili umelé portfólio z reálnych dát. Váhy portfólia nagenarujeme rovnako ako predtým, ako náhodnú prechádzku z normálneho rozdelenia $N(0, 0.005^2)$.

Pre úplnosť sme ešte pridali do regresie podmienku na súčet váh portfólia rovný 1. Nastavili sme rolovacie okno na šírku 50 a $\tau = 5000$. Výsledné odhady váh sú na obrázku 4.7. Obrázok hore ukazuje ako sme váhy nagenarovali, obrázok dole to, ako ich odhadla regresia.

Už z obrázku 4.7 vidieť, že s takýmto príkladom má použitá odhadovacia metóda problém a že odhady váh sa výrazne líšia od tých skutočných váh. Toto potvrdzujú aj hodnoty zrovnávacích kritérií napočítané v tabuľke 4.9.



Obr. 4.7: Váhy skutočné (hore) a vyfiltrované (dole) pre reálne portfólio s 8 zložkami pomocou rolovacej regresie.

súčet štvorcov	111.033
relatívny súčet štvorcov	0.0275379
počet nesprávnych odhadov	3835
relatívny počet nesprávnych odhadov	0.951141

Tabuľka 4.9: Hodnoty kritérií pre reálne portfólio s 8 zložkami odhadované regresiou.

5. Multikolinearita v dátach

Multikolinearita je štatistický jav, kedy sú dve alebo viac premenných v regresnom modeli vysoko lineárne korelované. V takejto situácii sa odhad parametrov môže nevyspytateľne meniť v reakcii na malú zmenu modelových dát. Kolinearita je lineárny vzťah medzi dvoma vysvetľujúcimi premennými. Dve premenné sú dokonale kolineárne, ak medzi nimi existuje presný lineárny vzťah. Napr. X_1 a X_2 sú perfektne kolineárne, ak existujú parametre λ_0 a λ_1 také, že pre každé pozorovanie i platí

$$X_{2i} = \lambda_0 + \lambda_1 X_{1i}.$$

K multikolinearite dochádza v prípade, ak dve alebo viac premenných v regresnom modeli je výrazne lineárne korelovaných. K dokonalej multikolinearite dochádza, ak napríklad v rovnici vyššie, je korelácia medzi dvoma nezávislými premennými rovná 1 alebo -1. V praxi sa však málokedy stretávame s dokonalou multikolinearitou dát. Oveľa častejšie dochádza k multikolinearite v prípadoch, kedy je medzi dvoma a viac premennými výrazný lineárny vzťah. Matematicky je množina premenných dokonale multikolineárna, ak existuje jeden alebo viac exaktných lineárnych vzťahov medzi premennými. Napríklad ak

$$\lambda_0 + \lambda_1 X_{1i} + \lambda_2 X_{2i} + \dots + \lambda_k X_{ki} = 0$$

platí pre každé i , kde λ_j sú konštanty a X_{ji} je i -te pozorovanie j -tej vysvetľujúcej premennej. Jedným z problémov, ktoré multikolinearita spôsobuje, je problém pri invertovaní matice $X^T X$ v algoritme regresie hľadajúcej odhady parametrov metódou najmenších štvorcov v rovnici

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \epsilon_i.$$

V prípade, že existuje dokonalá multikolinearita v dátach totiž matica $X^T X$ nemá plnú hodnotu a teda nie je invertibilná. V prípade, že nejde o dokonalú multikolinearitu, vyššie uvedená rovnica sa modifikuje na rovnicu s chybou μ_i

$$\lambda_0 + \lambda_1 X_{1i} + \lambda_2 X_{2i} + \dots + \lambda_k X_{ki} + \mu_i = 0.$$

V tomto prípade nemáme dokonalú multikolinearitu, ale premenné X_j sú takmer dokonale multikolineárne, ak rozptyl chýb μ_i je malý pre nejakú množinu hodnôt λ . V tomto prípade síce matica $X^T X$ má inverziu, ale výpočtový algoritmus ju nemusí byť schopný spočítať (aproximovať) a ak ju aj spočíta, táto môže byť veľmi citlivá aj na malé zmeny v dátach.

5.1 Detekcia multikolinearity

K detekovaniu multikolinearity v dátach existuje niekoľko metód. Hneď 6 spôsobov ako určiť multikolinearitu je spolu s vyššie uvedenou teóriou možné nájsť v [13]. Jedným zo spôsobov popísaných v [13] je vypočítanie variačného inflačného faktora (VIF), ktorý sa definuje ako

$$VIF = \frac{1}{1-R_j^2},$$

kde R_j^2 je koeficient determinace j-teho regresoru na ostatné regresory. Bežne používaná hranica pre rozhodnutie o probléme multikolinearity je VIF rovné 10. Hoci existuje článok [14], ktorý poukazuje na nevhodnosť tohto použitia pre určité prípady, pre naše účely budeme VIF rovný 10 považovať za threshold. Druhým kritériom, na základe ktorého môžeme predpokladať multikolinearitu, je vysoká korelácia medzi niektorými premennými. Vypočítaním výberovej korelačnej matice regresorov a porovnaním jednotlivých nediagonálnych prvkov v absolútnej hodnote s číslom 0.8. Formálne zapísané, ak

$$|r_{ij}| > 0.8$$

potom zrejme dvojica regresorov x_i a x_j spôsobuje multikolinearitu.

5.2 Multikolinearita v cenách akcií

V tejto práci využívame v niektorých príkladoch ceny (resp. z nich napočítané výnosy) akcií, ktoré sme stiahli zo stránok pražskej burzy. Ide o akcie významných spoločností, ktorých ceny na finančnom trhu nie sú nezávislé, z čoho by mohli vyplývať nepresnosti v odhadoch, ku ktorým dochádza práve pri príkladoch s 8 reálnymi zložkami. Overíme preto multikolinearitu týchto 8 akcií dvoma vyššie uvedenými spôsobmi. Najskôr napočítame korelačnú maticu cien týchto akcií.

$$\begin{pmatrix} 1. & 0.8978 & 0.8503 & 0.2085 & -0.6574 & 0.3001 & 0.418 & 0.8641 \\ 0.8978 & 1. & 0.8609 & 0.1487 & -0.5618 & 0.2815 & 0.2703 & 0.8258 \\ 0.8503 & 0.8609 & 1. & 0.2595 & -0.3562 & 0.4125 & 0.5117 & 0.7848 \\ 0.2085 & 0.1487 & 0.2595 & 1. & -0.3866 & -0.0756 & 0.2395 & -0.0117 \\ -0.6574 & -0.5618 & -0.3562 & -0.3866 & 1. & 0.1612 & -0.0124 & -0.4549 \\ 0.3001 & 0.2815 & 0.4125 & -0.0756 & 0.1612 & 1. & 0.3613 & 0.3341 \\ 0.418 & 0.2703 & 0.5117 & 0.2395 & -0.0124 & 0.3613 & 1. & 0.4636 \\ 0.8641 & 0.8258 & 0.7848 & -0.0117 & -0.4549 & 0.3341 & 0.4636 & 1. \end{pmatrix}$$

Vidíme, že pod diagonálou máme hneď 5 korelácií, ktoré prekračujú hranicu 0.8 a teda zrejme svedčia o multikolinearite.

Napočítaním variačného inflačného faktora pre regresory dostávame rovnaké výsledky. VIF vychádza

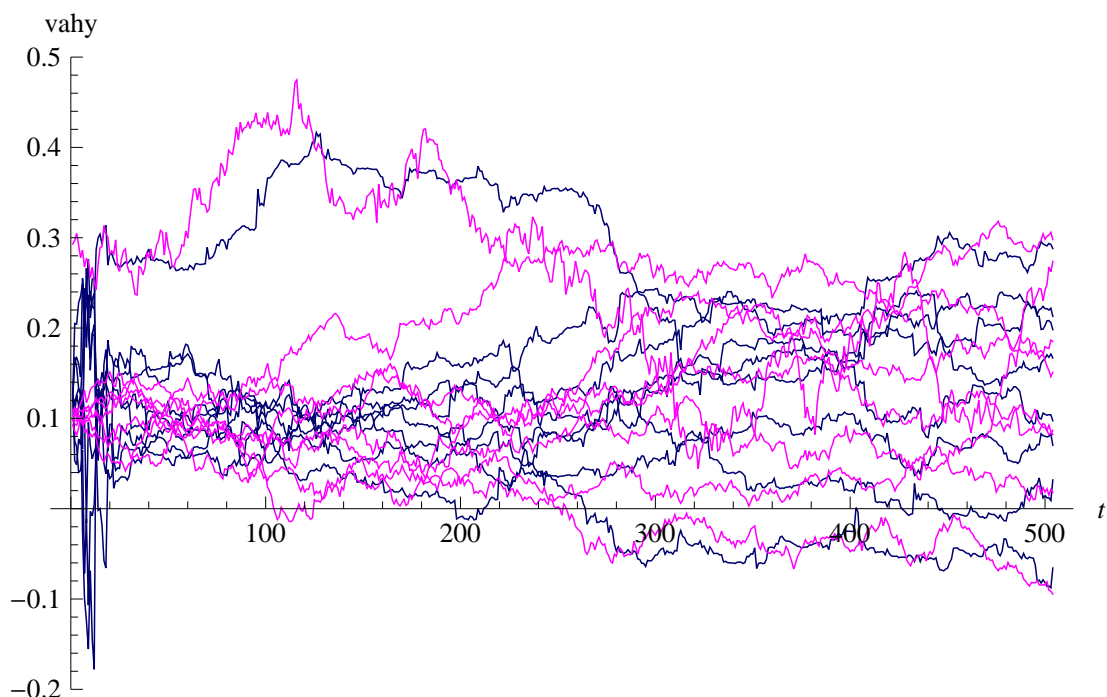
$$(13.22, 8.86, 8.51, 1.95, 4.18, 1.49, 2.16, 5.77)$$

Pre tri najhoršie vychádzajúce VIFy sme dostali hodnoty 13, 8.9 a 8.5, čo už sú hodnoty blízko hranice 10, resp. jedna nad. Na základe týchto výsledkov môžeme usudzovať o multikolinearite v použitých dátach. Táto nám zároveň dáva vysvetlenie pre nepresné odhady regresie a síce presnejšie, ale tiež oproti ideálnym prípadom horšie výsledky Kalmanovho filtra pre reálne dáta.

6. Porovnanie metód

6.1 Porovnanie na umelých a poloumelých dátach

Pripomeňme si, že rovnaký príklad, ako na konci kapitoly o rolovacej regresii, sme riešili aj pomocou Kalmanovho filtra. Práve pre možnosť porovnania sme použili rovnaké výnosy aj rovnaké zloženie portfólia ako v kapitole 4. Ešte raz si vykreslíme výsledky Kalmanovho filtra.



Obr. 6.1: Graf skutočných (ružové) a vyfiltrovaných váh pre portfólio s 8 reálnymi zložkami.

	Kalmanov filter	Rolovacia regresia
súčet štvorcov	7.01601	111.033
relatívny súčet štvorcov	0.00177531	0.0275379
počet nesprávnych odhadov	3222	3835
relatívny počet nesprávnych odhadov	0.799107	0.951141

Tabuľka 6.1: Hodnoty kritérií pre portfólio s 8 reálnymi zložkami pre obe metódy.

Porovnaním grafických výsledkov z obrázkov 4.7 a 6.1 a tiež porovnaním porovnávacích kritérií môžeme jednoducho dospieť k záveru, že Kalmanov filter

zvláda multikolinearitu dát oveľa lepšie ako rolovacia regresia. V niektorých prípadoch, ktoré sme v práci uviedli, však dopadla lepšie regresia.

Ako si teda vedie Kalmanov filter v porovnaní s rolovacou regresiou? Zhrňme výsledky niekoľkých príkladov zahŕňajúcich rôzne varianty, ktoré už sme v práci počítali a porovnajme úspešnosť metód.

Na začiatok upresnenie čo budeme porovnávať. V kapitole o rolovacej regresii sme našli nastavenie parametrov τ a šírky okna, ktoré sme prehlásili za najvhodnejšie z pohľadu minimalizovania štvorcových odchýliek. Je to najlepšie nastavenie rolovacej regresie aké sme našli, a preto ho budeme používať. Rolovacia regresia, o ktorej budeme pri porovnávaní hovoriť, má $\tau = 5000$ a *šírku okna* = 50 (pre váhy zložiek portfólia meniace sa náhodnou prechádzkou). Porovnanie budeme robiť na príkladoch bez zahrnutia Jensenovho alpha (pretože do rolovacej regresie sme ho nezahŕňali a u Kalmanovho filtra tvrdíme, že dáva lepšie výsledky bez neho).

Porovnanie metód uskutočníme na niekoľkých príkladoch. Pozrieme sa na príklad filtrácie s troma zložkami, ktoré sú generované z normálneho rozdelenia a umelo zložené do portfólia. Potom prejdeme k príkladu s troma zložkami reálnych výnosov stiahnutých zo stránok pražskej burzy, tiež umelo zloženými do portfólia. A potom sa pozrieme na obe varianty pre 8 zložiek portfólia t.j. 8 generovaných zložiek a 8 reálnych zložiek.

Pre všetky tieto alternatívy napočítame zavedené kritériá, súčet štvorcových odchýliek aj počet nepresne odhadnutých hodnôt. Výsledky tohto porovnania uvádzame v tabuľke 6.2.

počet zložiek	dáta	Kalmanov filter		Rol. Regresia	
		súčet štvorcov	počet "chýb"	súčet štvorcov	počet "chýb"
3	umelé	0.960204	1032	0.580167	963
3	reálne	2.13254	1224	2.36619	1280
8	umelé	2.45209	2709	10.3869	3341
8	reálne	7.01601	3222	111.033	3835

Tabuľka 6.2: Hodnotenie kritérií pre obe metódy.

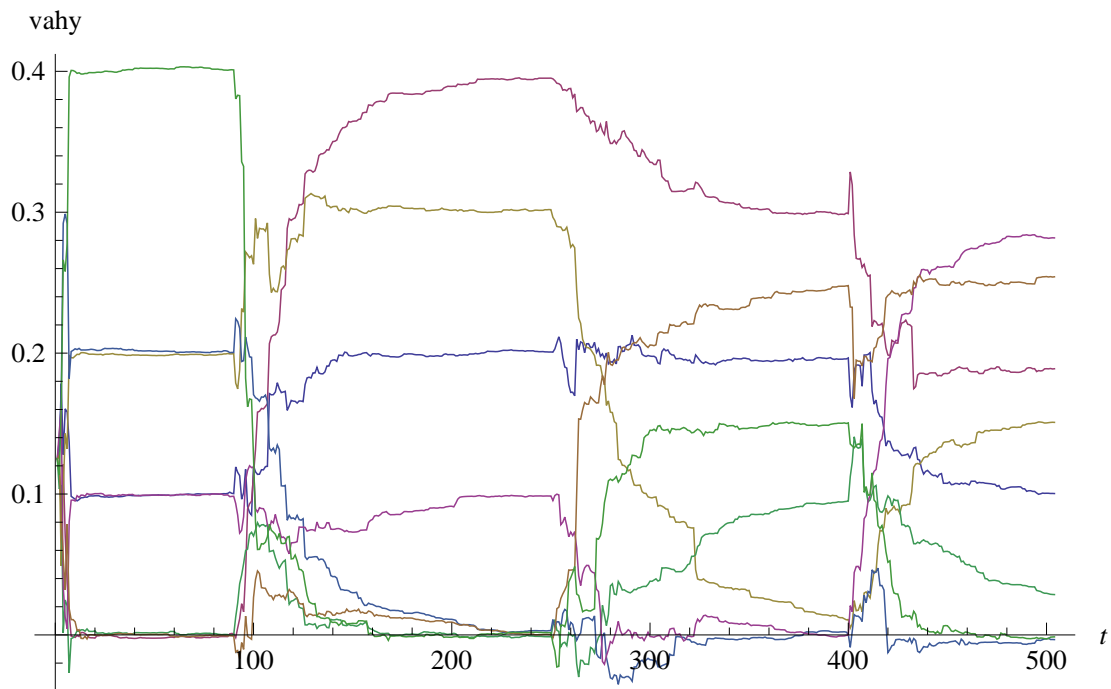
Porovnaním na základe kritéria súčtu štvorcov môžeme dospieť k záveru, že filtrácia pomocou Kalmanovho filtra dáva presnejšie výsledky ako filtrácia pomocou lokálne váženej regresie. Regresné odhady sú lepšie len pre jediný najjednoduchší príklad na umelých dátach a porovnateľné pre príklad s malým počtom reálnych zložiek. Toto môže byť ovplyvnené aj počtom pozorovaní, ktorý používame. V odbornnej literatúre sa pre regresné modely doporučuje mať minimálne 20 krát viac pozorovaní ako odhadovaných premenných, niekde dokonca hovoria o 30 až 50 násobku pozorovaní potrebnom pre stabilné výsledky. My v práci odhadujeme 8 parametrov a šírku okna máme nastavenú na 100 pozorovaní. V ideálnom prípade by sme potrebovali okno aspoň šírky 250, čo je pri reálnych dátach rok pozorovaní. Tým, že máme len pozorovania za dva roky (dĺžky 504), nastavením okna na polovicu z celkového počtu dát by model strácal požadovanú dynamiku.

Pozrieme sa ešte na jeden typ príkladu, kedy by mohlo byť užitočné využiť k dynamickej analýze rolovaciu regresiu. Ide o prípad, kedy by sme vedeli, že sa zmeny v portfóliu dejú skokovito, inak zostávajú váhy po častiach konštantné, prípadne len s malým šumom. V kapitole o rolovacej regresii sme našli nastavenie parametrov $\tau = 50$ a *sírku okna* = 20, ktoré sme prehlásili za ideálne pre prípad po častiach konštantných váh. Tie budeme v tomto príklade používať. Parametre tohto príkladu uvádzame v tabuľke 6.3.

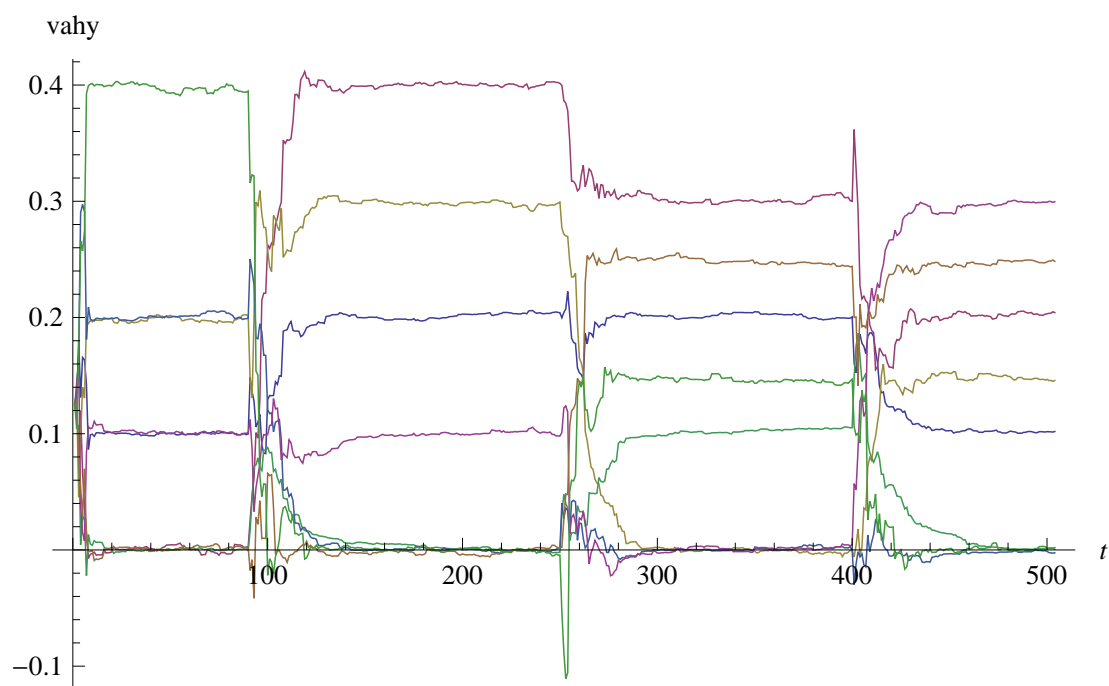
počet zložiek portfólia	8
rozdelenie náh. prechádzky	$N(0, 0.0001^2)$
rozdelenie pozorovacieho šumu	$N(0, 0.0001^2)$
Jensenovo alpha zahrnuté	áno/nie/nie
výnosy portfólia	reálne

Tabuľka 6.3: Parametre príkladu s po častiach konštantnými váhami so skokmi.

Tieto parametre sú platné pre tri varianty, ktoré budeme skúmať, preto tá informácia o zahrnutí Jensenovho alpha. Najskôr portfólio vyfiltrujeme pomocou Kalmanovho filtra, kde zahrnieme Jensenovo alpha, potom ešte raz bez Jensenovho alpha a nakoniec použijeme na rovnaký príklad ešte rolovaciu regresiu bez Jensenovho alpha. Pozrime sa na grafické výsledky týchto filtrácií.



Obr. 6.2: Graf vyfiltrovaných váh pre portfólio s po častiach konštantnými váhami so skokmi pomocou Kalmanovho filtra so zahrnutím Jensenovho alpha.

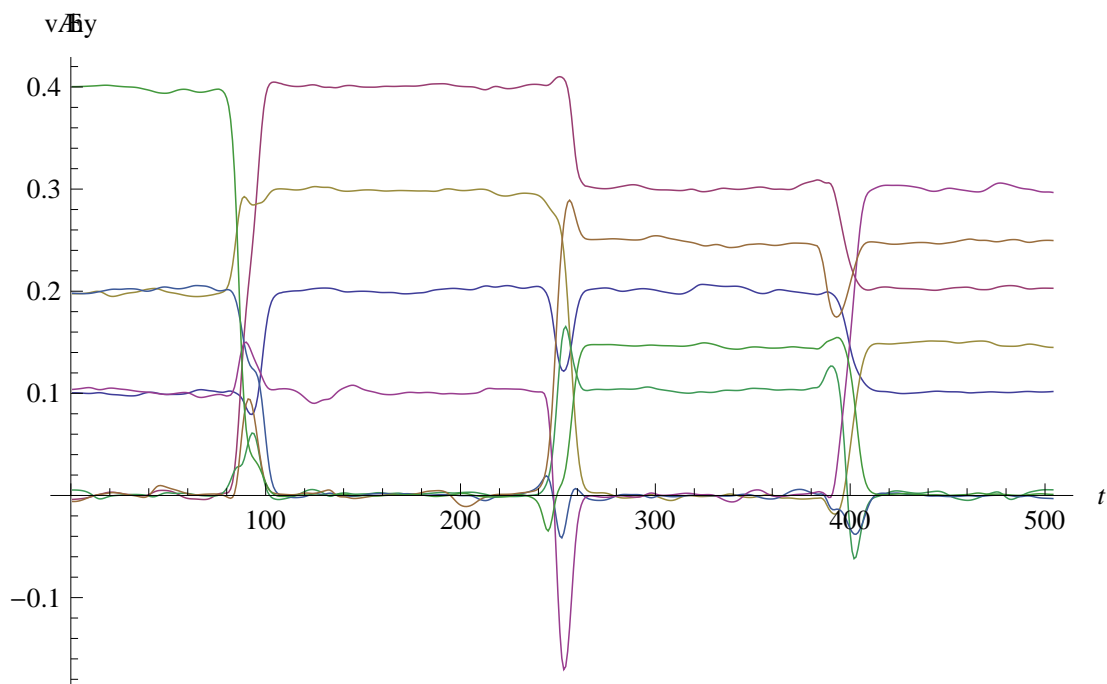


Obr. 6.3: Graf vyfiltrovaných váh pre portfólio s po častiach konštantnými váhami so skokmi pomocou Kalmanovho filtra bez Jensenovho alpha.

	KF s JA	KF bez JA	Rolovacia regresia
súčet štvorcov	11.5986	5.25568	2.3568
relatívny súčet štvorcov	0.00575325	0.00260699	0.00116905
počet nesprávnych odhadov	1678	675	410
relatívny počet nesprávnych odhadov	0.832341	0.334821	0.203373

Tabuľka 6.4: Hodnoty kritérií pre portfólio s lineárnymi váhami so skokmi.

Hodnoty kritérií pre všetky tri varianty uvádzame v tabuľke 6.4. Z tabuľky 6.4 vidieť, že v určitých prípadoch môže byť rolovacia regresia lepší filtračný nástroj aj pre reálne výnosy so silnou multikolinearitou. Zo všetkých troch použitých variant dáva rolovacia regresia najpresnejšie výsledky, čo je spôsobené najmä tým, že je ju možné dobre nastaviť a upraviť. V tomto prípade sme zúžili použitú šírku okna tak, aby pri skokovitej zmene váhy došlo k čo najrýchlejšej konvergencii k novej váhe. Tomu sme prisôbili aj hodnotu τ a dostali sme rolovaciu regresiu so šírkou okna 20 a $\tau = 50$ (optimálne nastavenie parametrov pre lineárne váhy so skokmi nájdené v kapitole o rolovacej regresii).



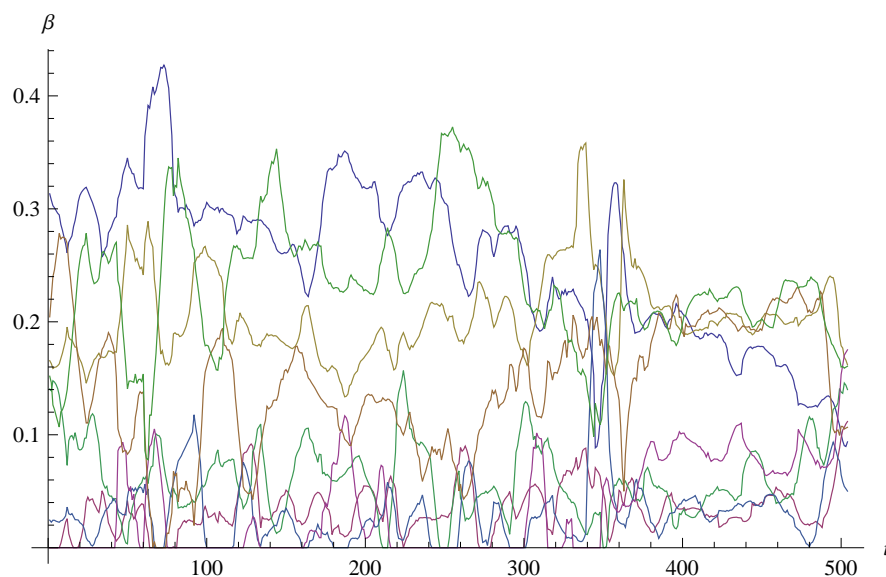
Obr. 6.4: Graf vyfiltrovaných váh pre portfólio s po častiach konštantnými váhami so skokmi pomocou rolovacej regresie.

6.2 Reálny príklad PX indexu

V predošlých kapitolách sme videli, že Kalmanov filter aj rolovacia regresia fungujú uspokojivo pre rôzne druhy umelých portfólií. Tieto boli zložené nielen z umelo generovaných dát, ale aj z reálnych dát, stiahnutých zo stránky pražskej burzy. V kapitole 3 uvádzame, že Kalmanov filter má určité nedostatky, ak príde na reálne dáta a reálne portfólio. Použijeme preto na rovnaký príklad rolovaciu regresiu a porovnáme obe metódy. Pri výpočte využijeme rovnakú myšlienku ako u Kalmanovho filtra o nevýznamnosti málo zastúpených akcií a preto úlohu rovno riešime pre 8 najvýznamnejších akcií PX indexu. Vyfiltrované váhy vidíme na obrázku 6.5.

Vidíme, že tým, že sme vynechali niektoré akcie z výpočtu, museli obe metódy vynechanú časť doplniť inde. Akcie, ktoré sme do výpočtu nezahrnuli, tvorili celkovo 2.6% portfólia a teda vyfiltrované váhy by mali nejak pokryť tento rozdiel. V skutočnosti ani Kalmanov filter ani rolovacia regresia nezvládli odhadnúť všetky váhy primerane. Kalmanov filter príliš podcenil váhu Vienna Insurance Group na úkor ČEZu a Komerční banky, zatiaľ čo rolovacia regresia precenila váhy New World Resources a Unipetrolu na úkor niekoľkých iných. Tieto nepresnosti v odhadoch sú zrejme spôsobené výraznou multikolinearitou.

Kalmanovmu filtru by ale v tomto prípade mohlo pomôcť prídanie podmienky na maximálnu možnú váhu. Táto podmienka je dostupná z informácií poskytovaných pražskou burzou a hovorí, že od marca roku 2011 nemôže váha jednej akcie v PX indexe prekročiť 20%. Tým by sa mohlo zabezpečiť, že filter neodhadne váhu ČEZu na 25% a následne nebude musieť odhadovať váhu Vienna Insurance Group o 10% menšiu ako v skutočnosti je. Rozšíreniami Kalmanovho filtra sa budeme zaoberať v poslednej kapitole o vylepšeniach metód.



Obr. 6.5: Váhy PX indexu pomocou rolovacej regresie.

Porovnajme výsledky Kalmanovho filtra a regresie. Váhy k 30.12.2011, reálne a vyfiltrované oboma metódami, uvádzame v tabuľke 6.5

Název	Váha [%]	Rolovacia regresia	Diff	Kalmanov filter	Diff
ČEZ	20,3	16,1%	-4,2%	25,0%	4,7%
ERSTE GROUP BANK	17,76	9,4%	-8,4%	19,2%	1,4%
KOMERČNÍ BANKA	16,92	16,2%	-0,7%	20,1%	3,2%
TELEFÓNICA C.R.	16,5	10,7%	-5,8%	17,3%	0,8%
VIG	13,69	11,2%	-2,5%	3,3%	-10,4%
NWR	4,8	17,5%	12,7%	4,4%	-0,4%
UNIPETROL	4,15	14,0%	9,9%	7,3%	3,1%
PHILIP MORRIS ČR	3,22	5,0%	1,8%	3,5%	0,3%
TOTAL (in abs values)			46%		24,3%

Tabuľka 6.5: Porovnanie výsledkov filtrácie.

6.3 Zhrnutie poznatkov

V predošlých kapitolách sme vyskúšali obe metódy filtrovania na rôzne druhy dát a pre rôzne prípady. Pokúsime sa teraz zhrnúť naše pozorovania a zosumarizovať v čom spočívajú výhody/nevýhody každej z metód. Najskôr si uvedieme tabuľku s akýmsi prehľadom výhod a nevýhod (6.6).

	Kalmanov filter	Rolovacia regresia
výhody	v mnohých softwaroch už predprogramované funkcie	jednoduchá ručná implementácia
	existuje veľké množstvo rozšírení	ľahko nastaviteľná pomocou parametrov
	lepší na reálne dáta	lepšia na umelé dáta
nevýhody	kladie predpoklady na vstupné dáta	vyžaduje dostatočné množstvo dát

Tabuľka 6.6: Zhrnutie porovnania metód.

Ako uvádza tabuľka 6.6, výhodou Kalmanovho filtra je, že ho človek nemusí ručne implementovať, ale je v rôznych matematických softwaroch už predprogramovaný ako funkcia (napr. Eviews, Mathematica). Toto samozrejme môžeme považovať za výhodu, na druhej strane, ak chceme Kalmanov filter a to ako funguje pochopiť a nie len pracovať s „blackboxom“, strávime nakoniec veľa času študovaním týchto funkcií, ktoré sa zároveň len ťažko upravujú. Oproti tomu rolovacia regresia je jednoduchá na ručnú implementáciu a na prvý pohľad priehľadná.

Ďalšou výhodou Kalmanovho filtra sú jeho rozšírenia. Existuje množstvo článkov zaoberajúcich sa pridávaním podmienok do Kalmanovho filtra, rozširovaním filtra na nelineárne modely a pod. Rolovacia regresia je zase ľahko prispôsobiteľná. Ak sa človek rozhodne pre lokálne váženú regresiu, vie si ju jednoducho nastaviť využitím dvoch parametrov a rovnako ľahko sa do nej pridávajú podmienky (avšak aspoň v Mathematice výrazne spomaľujú výpočet).

Výhoda Kalmanovho filtra tiež spočíva vo filtrácii na reálnych dátach. Ako sme videli vo viacerých prípadoch, Kalmanov filter si oveľa lepšie poradí s komplikáciami ako je korelácia dát a multikolinearita. Regresia však dáva presnejšie výsledky pre umelé dáta a je ľahko nastaviteľná, aby rýchlo indikovala skoky.

Posledná výhoda regresie spočíva v predpokladoch. Zatiaľ čo Kalmanov filter kladie predpoklady na vstupné dáta, v regresii ich nemáme. Tu je namieste poznámka z [9]. V článku sa uvádza, že Kalmanov filter dáva najlepší odhad pre dynamické lineárne systémy, ak je splnená podmienka na gaussovský šum a zároveň dáva najlepší lineárny odhad pre dynamické lineárne systémy, ak podmienka nie je splnená.

7. Vylepšenie metód

Na záver práce si ešte ukážeme spôsob, ako by šlo obe metódy filtrácie vylepšiť. Možností a spôsobov je niekoľko, my sa zameriame na dva spôsoby.

Prvý spôsob spočíva v pridávaní podmienok. Najmä na Kalmanov filter existuje niekoľko metód, popísaných v rôznych článkoch, ako pridať podmienky pre filtrované váhy. Veľa z nich je popísaných napríklad v článku [9].

Druhý spôsob spočíva vo využití predpokladu, že zmeny zloženia portfólia sa dejú skokovito. Zahnutím tohoto poznatku by šli obe doteraz spomenuté metódy ešte spresniť. Myšlienka je jednoduchá. Vieme, že zmeny v našom portfóliu sú skokové a teda že na intervale, kde regresia konverguje k zmene a kde Kalmanov filter mení hodnoty je jedna jediná zmena. Vieme tiež (máme vyfiltrované), aká bola príslušná váha pred zmenou, aj aká je po zmene, nevieme ale, kde zmena nastala. A práve k riešeniu tohoto problému - nájdenia miesta skoku - slúžia štatistické testy nazývané CUSUM testy.

7.1 Rozšírenie Kalmanovho filtra

Kalmanov filter je rozšírená a v poslednej dobe často skúmaná metóda, ku ktorej existuje množstvo článkov o vylepšení. My sa zameriame na rozšírenie o pridanie podmienok, ktorým sa zaoberá článok [9]. Tu autor uvádza niekoľko spôsobov ako pridať podmienky do filtra. Najskôr si ukážeme ako pridať podmienku na rovnosť v tvare

$$D\hat{x}_t = d.$$

Pripomeňme si rovnice lineárneho systému, s ktorými pracujeme:

$$y_t = G_t x_t + w_t,$$

$$x_t = F_t x_{t-1} + v_t.$$

Predstavme si teraz, že chceme do systému zahrnúť podmienku

$$\begin{pmatrix} 1 & 0 & 1 \end{pmatrix} x_t = 0.$$

Toto je ekvivalentné podmienke $x_t[1] = -x_t[3]$, kde hranaté zátvorky označujú zložku vektora x_t . Postup pri zapracovaní tejto podmienky je jednoduchý. Tým, že si vieme vyjadriť $x_t[3]$ pomocou $x_t[1]$, nepotrebujeme $x_t[3]$ v rovnici. Preto ho vynecháme, čím znížime dymenziu sústavy o jedna. Tento postup sa v [9] nazýva „redukcia modelu“.

Ďalším rozšírením je v kapitole 3 už spomínané „Perfect measurement“, ktoré sme využili pri pridaní podmienky na súčet váh rovný 1. Ak sú podmienky opäť dané rovnosťou $D\hat{x}_t = d$, kde D je matica typu $s \times n$, kde $s \leq n$, môžeme podmienku $D\hat{x}_t = d$ zahrnúť do sústavy nasledovne:

$$\begin{pmatrix} y_t \\ d \end{pmatrix} = \begin{pmatrix} G_t \\ D \end{pmatrix} x_t + \begin{pmatrix} w_t \\ 0 \end{pmatrix}.$$

Stavová rovnica lineárneho systému zostala nezmenená, a rovnicu pozorovaní sme rozšírili. Tento prístup je ekvivalentný prístupu redukcii modelu.

7.1.1 Projekcia odhadov

Ďalším prístupom k filtrovaniu s obmedzeniami je projekcia nepodmienených odhadov \hat{x}_t^+ Kalmanovho filtra do podpriestoru obmedzení. Odhad s obmedzením sa potom dá zapísať ako

$$\bar{x}_t^+ = \operatorname{argmin}_x (x - \hat{x}_t^+)' W (x - \hat{x}_t^+)$$

tak, že $Dx = d$, kde W je pozitívne-definitná váhová matica. Riešením je potom

$$\bar{x}_t^+ = \hat{x}_t^+ - W^{-1} D' (D W^{-1} D')^{-1} (D \hat{x}_t^+ - d).$$

Ak proces aj šumy sú gaussovske a nastavíme $W = (P_t^+)^{-1}$ dostaneme maximálne vierohodný odhad stavového vektora vzhľadom k obmedzeniam. Ak nastavíme $W = I$ dostaneme odhad metódou najmenších štvorcov stavového vektora vzhľadom k obmedzeniam. Dá sa ukázať, že odhad s obmedzením \hat{x}_t^+ je nestraný. Teda

$$E(\bar{x}_t^+) = E(x_t).$$

Nastavením $W = (P_t^+)^{-1}$ dostaneme filter s minimálnym rozptylom, t.j.

$$\operatorname{Cov}(x_t - \bar{x}_t^+) \leq \operatorname{Cov}(x_t - \hat{x}_t^+)$$

pre každé \hat{x}_t^+ . Nastavením $W = I$ dostávame odhad s obmedzením, ktorý je bližšie k skutočnému stavu ako odhad bez obmedzenia a to v každom kroku. T.j.

$$\|x_t - \bar{x}_t^+\|_2 \leq \|x_t - \hat{x}_t^+\|_2$$

pre všetky t .

Tento prístup projekcie odhadov do podpriestoru obmedzení má výhodu, že sa dá rozšíriť aj na obmedzenia zadané nerovnosťou vyriešením

$$\bar{x}_t^+ = \operatorname{argmin}_x (x - \hat{x}_t^+)' W (x - \hat{x}_t^+)$$

s podmienkou

$$Dx \leq d.$$

Táto úloha je úlohou kvadratického programovania. Na riešenie takejto úlohy existuje niekoľko prístupov vrátane prístupu vnútorného bodu a metód aktívnych množín. Metóda aktívnych množín využíva fakt, že len obmedzenia, ktoré sú aktívne pri riešení problému, sú významné pre podmienky optimality. Predpokladajme, že máme s nerovnostných obmedzení a q z týchto s obmedzení sú aktívne pri riešení

$$\bar{x}_t^+ = \operatorname{argmin}_x (x - \hat{x}_t^+)' W (x - \hat{x}_t^+),$$

$$Dx \leq d.$$

Označme \hat{D} q riadkov D , ktoré zodpovedajú aktívnym obmedzeniam a označme \hat{d} tých q zložiek d , ktoré tiež zodpovedajú aktívnym obmedzeniam. Ak poznáme množinu aktívnych obmedzení *a priori* potom riešením

$$\bar{x}_t^+ = \operatorname{argmin}_x (x - \hat{x}_t^+)' W (x - \hat{x}_t^+)$$

$$Dx \leq d$$

je tiež riešenie s obmedzeniami zadanými rovnosťami

$$\bar{x}_t^+ = \operatorname{argmin}_x (x - \hat{x}_t^+)' W (x - \hat{x}_t^+)$$

tak, že

$$\hat{D}x = \hat{d}.$$

Úloha s obmedzeniami pomocou nerovností je ekvivalentná úlohe s obmedzeniami pomocou rovností. Preto aj všetky vlastnosti odhadu s obmedzeniami danými rovnosťami platia aj pre odhady s obmedzeniami danými nerovnosťami.

Pripomeňme si teraz príklad s PX indexom v kapitole tri. Spomínali sme, že Kalmanov filter nadhodnotil podiel akcie ČEZu nad limit daný pražskou burzou. Rozšírime preto Kalmanov filter o podmienku na podiel ČEZu maximálne 20%. Teraz porovnáme výsledky pôvodného a rozšíreného Kalmanovho filtra v tabuľke 7.1

Názov	Váha [%]	Kalmanov filter	Diff	KF s podmienkou	Diff
ČEZ	20,3	25,0%	4,7	20,0%	-0,3%
ERSTE GROUP BANK	17,76	19,2%	1,44	19,9%	2,1%
KOMERČNÍ BANKA	16,92	20,1%	3,18	20,8%	3,9%
TELEFÓNICA C.R.	16,5	17,3%	0,8	18,0%	1,5%
VIG	13,69	3,3%	-10,39	4,0%	-9,7%
NWR	4,8	4,4%	-0,4	5,1%	0,3%
UNIPETROL	4,15	7,3%	3,15	8,0%	3,8%
PHILIP MORRIS ČR	3,22	3,5%	0,28	4,2%	1,0%
TOTAL (in abs values)			24,34%		22,6%

Tabuľka 7.1: Porovnanie výsledkov rozšíreného Kalmanovho filtra.

Stĺpec označený ako Kalmanov filter s podmienkou, sú výsledky rozšírenej verzie Kalmanovho filtra, kde sú zaradené dve podmienky. Prvá, ktorú sme mali po celý čas, a to na súčet váh rovný 1 a druhá, ktorá nám zabezpečila, že váha pre ČEZ neprekročí 20%. Táto podmienka bola zadaná len pre túto jednu akciu, čím nám vyskočila pre zmenu akcia Komerční banky vyššie, nad 20%. Podmienku na maximálne 20% nú váhu udáva pražská burza, ale ako vidieť u skutočných čísel z konca roku 2011 mierne prekročenie sa zrejme toleruje.

Napočítané difference sú medzi výsledkami Kalmanovho filtra s podmienkou a reálnymi váhami k 30.12.2011.

Pri tejto projekcii sme použili jednotkovú maticu ako váhovú maticu projekcie ($W = I$). Jednotková váhová matica viedla k úprave odhadnutých dát jednoduchým, ale nie úplne najlepším spôsobom. Metóda projekcia spočívala v tom, že sa požadovaná váha znížila na nami zadanú hodnotu a rozdiel medzi týmito hodnotami sa rovnomerne rozdelil medzi ostatné váhy.

Pozriem sa preto aj na druhú možnosť projekcie odhadov. Miesto jednotkovej váhovej matice použijeme projekčnú maticu $W = (P_t^+)^{-1}$. Túto nám spolu s váhami odhadol Kalmanov filter. Váhová matica $(P_t^+)^{-1}$ zabezpečí, že rozdiel, o ktorý znížime jednu zložku portfólia, sa nerovnomerne rozloží medzi ostatné váhy. Výsledky projekcie sú v tabuľke 7.2.

Názov	Váha [%]	Kalmanov filter	Diff	KF s podmienkou	Diff
ČEZ	20,3	25,0%	4,7	20,0%	-0,3%
ERSTE GROUP BANK	17,76	19,2%	1,44	19,6%	1,8%
KOMERČNÍ BANKA	16,92	20,1%	3,18	20,7%	3,7%
TELEFÓNICA C.R.	16,5	17,3%	0,8	19,0%	2,5%
VIG	13,69	3,3%	-10,39	3,3%	-10,4%
NWR	4,8	4,4%	-0,4	5,3%	0,5%
UNIPETROL	4,15	7,3%	3,15	8,0%	3,8%
PHILIP MORRIS ČR	3,22	3,5%	0,28	4,1%	0,9%
TOTAL (in abs values)			24,34%		23,9%

Tabuľka 7.2: Porovnanie výsledkov rozšíreného Kalmanovho filtra s proporcionálnou projekciou.

V porovnaní s projekciou odhadov s jednotkovou váhovou maticou sa tri váhy zlepšili, tri zhoršili a dve sa nezmenili. Ponechávame preto na posúdenie čitateľa, ktorú z metód projekcie považuje za lepšiu.

7.1.2 CUSUM Kalman filter

Nakoniec našej práce si priblížime ešte jedno rozšírenie Kalmanovho filtra. V prípade, že by sme predpokladali, že zloženie portfólia je po častiach konštantné, t.j. predpokladáme, že zloženie je po nejaký čas stabilné a potom dôjde ku skokovitej zmene, môžeme odhady váh portfólia ešte upresniť. Upresnenie spočíva v tom, že nájdeme, kedy skoky nastali a prehlásime váhy pred skokom a po ňom za konštantné.

Testov na zisťovanie takéhoto porušenia stability modelu je niekoľko. Rozšírenou metódou ako testovať stabilitu modelu je štatistický prístup, ktorý v praxi často vyžíva tzv. change detection testy, resp. CUSUM testy. Tieto testy umožňujú rozpoznať, resp. potvrdiť, existenciu dátových segmentov, ktorým v príslušnom modeli zodpovedajú odlišné parametrické hodnoty. Využitím poznatkov o nestabilite parametrov môžeme modely vylepšiť napr. tým, že odhadneme každý segment zvlášť alebo pridaním vysvetľujúcich premenných do jednotlivých segmentov.

Pridaním CUSUM testu (CUSUM odvodené od kumulatívnych súčtov) priamo do Kalmanovho filtra sa zaoberá článok [10].

Autori tu pracujú s takto rozšírenými rovnicami modelu:

$$y_t = G_t x_t + w_t,$$

$$x_t = x_{t-1} + \delta_t v_t,$$

kde δ_t je buď 0 alebo 1 podľa toho, či nastal skok. Takto upravený model odstraňuje malé šumy z váh a váhy x_t mení len pri veľkých skokoch.

Filtrácia potom prebieha rekurzívnymi výpočtami s využitím rovníc

$$\hat{x}_{t+1|t} = \hat{x}_{t|t-1} + K_t(y_t - G_t \hat{x}_{t|t-1}),$$

$$K_t = P_{t|t-1} G_t' (G_t P_{t|t-1} G_t' + R_t)^{-1},$$

$$P_{t+1|t} = [P_{t|t-1} - P_{t|t-1} G_t' (G_t P_{t|t-1} G_t' + R_t)^{-1} G_t P_{t|t-1}] + Q_t,$$

ktoré zodpovedajú rekurzívnym vzorcom z druhej kapitoly (s menšími úpravami - odstránili sme maticu F_t , ktorú tu nepoužívame a v skutočnosti sme ju mali celý čas nastavenú ako jednotkovú).

To, v čom spočíva zmena je zahrnutie testovacej štatistiky do rekurzívnych výpočtov. V každom kroku filtrácie sa okrem vyššie uvedených vzorcov ešte napočítava:

$$s_t = y_t - G_t \hat{x}_{t-1}$$

$$g_t = \max(g_{t-1} + s_t - \xi, 0)$$

Testová štatistika g_t sa porovnáva s hodnotou h a ak platí

$$g_t > h,$$

test zahlási zmenu, δ_t sa nastaví rovné 1, skok sa zahrnie do váh a následne sa všetky hodnoty nastaví na počiatočné, t.j. $\delta_t = 0$, $g_t = 0$. A opäť sa rekurzívne napočítava štatistika.

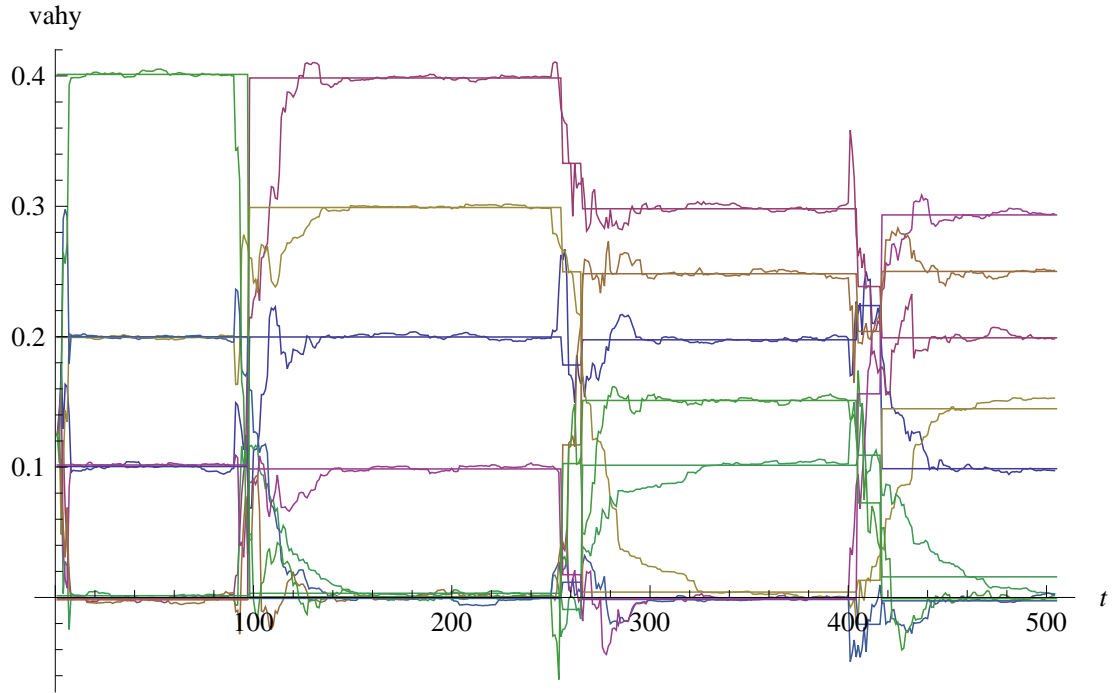
CUSUM test má dva stupne voľnosti. Záporný posun ξ a threshold pre zaznamenanie skoku h . Tieto hodnoty sa určujú experimentálne.

Postup, ako ξ a h nastaviť je v [11] popísaný nasledovne:

1. Začnime s veľkým thresholdom h .
2. Nastavme ξ na polovicu očakávanej zmeny, alebo nastavme ξ tak, aby $g_t = 0$ pre viac ako 50% prípadov.
3. Prestavme threshold tak, aby sme dostali požadovaný počet falošných alarmov alebo požadované omeškanie pre detekciu zmeny.

Ďalšie možné úpravy nastavení parametrov sa odvíjajú od požiadaviek

- ak požadujeme rýchlejšiu detekciu zmien, skúsme zmenšiť ξ
- ak požadujeme menej falošných alarmov, zväčšime ξ



Obr. 7.1: Graf vyfiltrovaných váh a ich ekvivalentov odhadnutých CUSUM testom.

Takto uvedený test detekcie zmien je jednostranný. Test predpokladá, že všetky s_t sú kladné čísla. V našom prípade však pripúšťame skoky váh zložiek portfólia oboma smermi, preto by sme mali vyššie uvedený test rozšíriť podľa [11] na obojstranný. Rozšírenie spočíva v tom, že sa v každom kroku budú spočítavať dve testové štatistiky g_{t1} a g_{t2} , kde

$$g_{t1} = \max(g_{t1-1} + s_t - \xi, 0),$$

$$g_{t2} = \min(g_{t2-1} + s_t + \xi, 0).$$

Testové štatistiky g_{t1} a g_{t2} sa porovnávajú s hodnotou h a ak platí

$$g_{t1} > h$$

alebo

$$g_{t2} < -h$$

test zahlási zmenu, δ_t sa nastaví rovné 1, skok sa zahrnie do váh a v ďalšom kroku sa všetky hodnoty nastaví na počiatočné, t.j. $\delta_t = 0$, $g_{t1} = 0$ a $g_{t2} = 0$. A opäť sa rekurzívne napočítavajú štatistiky.

Výstupom CUSUM testu môžu byť priamo váhy, ak by sme pridali testovanie do rekurzívneho výpočtu a použili δ_t tak ako sa uvádza v článku [10], t.j. ako indikátor toho, či je detekovaná dostatočne veľká na to, aby sme ju pridali. Pri Kalmanovom filtri, tak ako ho máme naprogramovaný, však narazíme na problém v tom, že ak test identifikuje nejaký skok, nemusí to byť ten správny, keďže filter v rámci konverencie mení váhy skokovito. Riešením tohto problému je využiť CUSUM test len k nájdeniu skokov a veľkosť skoku určiť inak. Výhodou tohto

postupu je aj to, že ho môžeme aplikovať na už vyfiltrované váhy a nemusíme testovanie pridávať do rekurzívnych výpočtov. Tým dostaneme intervaly, o ktorých na základe predpokladov vieme, že váhy sú tu konštantné. Váhy potom stanovíme ako medián váh odhadnutých Kalmanovým filtrom na danom intervale (prípadne opäť spustíme regresiu a váhy odhadneme regresiou).

Pozrime sa teraz na samotný test a jeho výsledky. Test aplikujeme na príklad s portfóliom s ôsmimi reálnymi výnosmi umelo zloženými do portfólia. Parametre ξ a h sme stanovili podľa vyššie uvedených inštrukcií experimentálne na $\xi = 0.00000112$ a $h = 0.002$. Dostali sme skoky v týchto časoch: 1, 2, 90, 241, 421. Nájsené skoky v časoch 1 a 2 môžeme interpretovať ako dôsledok počiatočnej konvergenzie, preto tu zmenu uvažovať nebudeme. Zostali nám skoky v časoch 90, 241 a 421. Skutočné skoky boli v časoch 90, 250 a 400. Na obrázku 7.1 sú vykreslené pôvodné vyfiltrované váhy portfólia a tiež po častiach konštantné váhy ako výsledok po aplikovaní CUSUM testu.

Vidíme, že tam, kde Kalmanov filter konverguje pomaly k správnym váham ich už rozšírená verzia po aplikácii CUSUM testu správne odhadla. Trochu menej presné výsledky sme dostali až na poslednom časovom intervale (400 až 504 dní), kde váha jednej akcie pomaly klesá k nule. Klesanie je tak pomalé, že použitá metóda mediánu na danom intervale neodhadla váhu správne na nulu, ale na 0.015 (práve tu by mohlo výsledkom pomôcť použitie regresie na jednotlivé konštantné intervaly). Napriek tejto nepresnosti môžeme CUSUM test, pridaný do Kalmanovho filtra, považovať za vylepšenie Kalmanovho filtra, ktoré v určitých situáciách spresní výsledky.

Záver

Cieľom tejto diplomovej práce bolo podrobne popísať problém dynamickej analýzy portfólia na základe jeho pozorovaných výnosov, analyzovať a interpretovať výsledky a navrhnúť ich vhodné vylepšenia.

V práci sme popísali Kalmanov filter a lokálne váženú regresiu ako dve metódy používané k dynamickej analýze portfólia. Uviedli sme potrebnú teóriu a odkazy na použitú literatúru pre možnosť rozšírenia informácií čitateľa.

V praktických aplikáciách pre Kalmanov filter sme najskôr popísali využitie softwaru Mathematica 8 a jeho balíčka TimeSeries pre naše účely. Uviedli sme vhodné nastavenie využívanej minimalizačnej funkcie FindMinimum, ktoré je nevyhnutné pre správne výsledky, keďže gradientná metóda minimalizácie sa pri nesprávnom nastavení môže zaseknúť na lokálnom minime.

V ďalšej časti sme skúmali význam predpokladov Kalmanovho filtra a výsledky filtra pre prípad, že predpoklady nie sú splnené. Ďalej sme uviedli praktické aplikácie filtra a poukázali na to, že pri reálnych dátach zafažených výraznou multikolinearitou klesá úspešnosť filtrácie (v zmysle zavedených porovnávacích kritérií a v porovnaní s dátami, u ktorých sú výnosy jednotlivých akcií nekorelované).

Pri dynamickej analýze pomocou lokálne vázenej regresie sme najskôr hľadali vhodné nastavenie parametrov pre rôzne druhy vstupných dát. Uviedli sme, ako správne nastaviť šírku rolovacieho okna a váhy pozorovaní v okne, k docieleniu čo najpresnejších výsledkov. Rovnako ako pre Kalmanov filter, sme uviedli praktické aplikácie na rôznych druhoch dát.

Praktické využitie oboch metód sme ilustrovali na príklade PX indexu. Zo stránky pražskej burzy sme si stiahli informácie o zložkách PX indexu (ceny akcií, z ktorých sme napočítali výnosy) aj výnosy samotného indexu a odhadovali sme jeho zloženie. Výsledky analýz oboch metód sme interpretovali pomocou porovnávacích kritérií a následne porovnali. Pri tomto porovnaní sme poukázali na silné a slabé stránky jednotlivých metód v rôznych situáciách, napr. jednoduchosť ručnej implementácie rolovacej regresie a jej ľahká nastaviteľnosť pomocou parametrov a proti tomu nenáročnosť Kalmanovho filtra na dáta (lepšie zvláda výrazne korelované dáta).

V závere práce sme navrhli modifikácie existujúcich metód tak, aby sme dosiahli čo najlepšie výsledky z hľadiska porovnania odhadovaného a skutočného vývoja zloženia portfólia v čase. Uviedli sme rozšírenie Kalmanovho filtra pomocou projekcie odhadov a ilustrovali sme možnosť pridania testu detekcie zmien (CUSUM test) do filtrácie.

Zoznam použitej literatúry

- [1] Králová D.: *Dynamická analýza portfolia pomocí Kalmanova filtru*, diplomová práce, Praha, 2010
- [2] Cipra T., Romera R.: *Robust Kalman filter and its applications in time series*, Kybernetika 27, 1991
- [3] Wolfram Mathematica: *Documentation center*, [Internet],
<http://media.wolfram.com/documents/TimeSeriesDocumentation.pdf>
- [4] Anderson B.D.O., Moore J.B.: *Optimal filtering*, Prentice-Hall, 1979
- [5] Harvey A.C.: *Forecasting, structural time series models and the Kalman filter*, Cambridge University press, 1989
- [6] Welch G., Bishop G.: *An Introduction to the Kalman Filter*, SIGGRAPH, 2001
- [7] Zivot E., Wang J.: *Modeling Financial Time Series with S-Plus*, Softcover, 2006
- [8] Fabozzi F.J., Gupta F., Markowitz H.M.: *The Legacy of Modern Portfolio Theory*, The Journal of Investing, 2002
- [9] Simon D.: *Kalman Filtering with State Constraints: A Survey of Linear and Nonlinear Algorithms*, Cleveland State University, 2009
- [10] Jacobsson K., Möller N., Johansson K.H., Hjalmarsson H.: *Some Modeling and Estimation Issues in Control of Heterogeneous Networks*, Department of Signals, Sensors and Systems, Royal Institute of Technology, Sweden, 2004
- [11] Wiley: *Adaptive filtering and change detection*, West Sussex, 2000
- [12] Dupač V., Hušková M.: *Pravděpodobnost a matematická statistika*, Nakladatelství Karolinum, 2005
- [13] Wikipedie [Internet]
<http://en.wikipedia.org/wiki/Multicollinearity>
- [14] O'Brien, Robert M.: *A Caution Regarding Rules of Thumb for Variance Inflation Factors*, Quality and Quantity, 2007

Prílohy